An Evaluation of the

Quality of Surgical Training

A Global Study

Pritam Singh MBBS MA(Cantab) MRCS(Eng)

Division of Surgery

Department of Surgery & Cancer

Imperial College London

Thesis submitted for the Degree of

Doctor of Philosophy (PhD)

2015

Supervisors:

Professor the Lord Ara Darzi MD FRS FMedSci FRCS FACS

Mr Rajesh Aggarwal MBBS PhD MA FRCS FRCSC FACS
Abstract

Quality assessment in healthcare is an essential part of governance culture. Clinical outcomes and research output are routinely assessed with national and global comparisons commonplace in a bid to continually drive up standards. While the subject of surgical training has received greater exposure in recent years, there remains no rigorous assessment of surgical training quality in the same way as clinical and research outcomes. To improve surgical training quality, high quality training must first be defined to permit evaluation of training and facilitate identification of methods to raise standards.

Variability of global surgical training programs is first identified by comparing the national surgical curricula of a sample of countries against defined standards for curricula. A questionnaire study further explores the variations between the two nations with the most contrasting surgical curricula.

Qualitative methodologies, including semi-structured interviews and a global Delphi consensus study, are subsequently employed to define high quality surgical training. The results have enabled the development of a surgical training quality assessment tool (S-QAT). The S-QAT has been piloted in the North West London training region where variability in surgical training quality has been confirmed.
Defining high quality training has facilitated the development and piloting of interventions to improve surgical training quality. The first intervention employs video-based coaching to improve surgical skills. The second intervention is an intensive, immersive week of simulation training to ease the transition from medical student to Foundation Doctor. These interventions demonstrate that improvements in training quality can be made without the need to rely on increasing the quantity or time in training.

The foundations for robust evaluation of surgical training quality have been constructed. Evidence-based medicine is the gold standard for patients; evidence-based training of surgeons will ensure standards are maintained for the next generation of patients.
Acknowledgements

I would firstly like to thank my supervisors. Professor Darzi gave me the prestigious opportunity to join the Academic Surgical Unit at Imperial College London. He has provided guidance and taught me much about leadership. Rajesh Aggarwal has been an excellent mentor and his perceptive feedback provided the ideal balance of direction while allowing me to retain a sense of autonomy over my research.

A highlight of my study has been the friendships and collaborations I have developed with my colleagues. Some have formally collaborated with me and their contributions are described later, I am extremely grateful for their efforts. In addition I have benefited from frequent informal conversations with researchers and support staff, these are not easy to quantify but have been very valuable. I would also like to thank all participants, not only for their generosity with their time, but also for their enthusiasm that fuelled my own motivation to study.

My family’s support has given me great strength throughout this endeavour. In particular, my mother’s unflinching and absolute confidence in my ability has always encouraged me in my studies. My wife Jagdeep has added to this with her patience, emotional support but most of all by ensuring that despite the demands of work, I have always remained balanced and happy. Lastly, I would like to thank my son Arvind whose natural exploration of all things around him has been an inspiration for me during my research.
Declaration

Originality

I declare that the studies described in this thesis were performed by myself under the supervision of Professor Ara Darzi and Mr Rajesh Aggarwal. This thesis has been composed by myself and any contributions by others have been appropriately referenced and any collaborations are appropriately credited.

Pritam Singh MBBS MA(Cantab) MRCS(Eng)

Copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Pritam Singh MBBS MA(Cantab) MRCS(Eng)
Table of Contents

Abstract ............................................................................................................................................ 2

Acknowledgements .......................................................................................................................... 4

Declaration ......................................................................................................................................... 5

Originality ........................................................................................................................................... 5

Copyright ........................................................................................................................................... 5

Table of Contents ............................................................................................................................ 6

List of Tables and Figures .................................................................................................................. 12

Contributions ..................................................................................................................................... 15

Research Outputs ............................................................................................................................. 17

Publications ......................................................................................................................................... 17

Presentations ....................................................................................................................................... 19

Chapter 1: Introduction ..................................................................................................................... 23

1.1: Aims of the Thesis ..................................................................................................................... 24

1.2: Overview & Rationale .................................................................................................................. 25

1.3: The Evolution of Surgical Training .......................................................................................... 30

1.3.1: Calman Reforms .................................................................................................................... 33

1.3.2: The ‘Lost Tribe’ ...................................................................................................................... 34

1.3.3: Modernising Medical Careers .............................................................................................. 38

1.3.4: Implementation of MMC reforms ....................................................................................... 42
1.3.5: Aspiring to Excellence ................................................................. 43
1.3.5: NHS Next Stage Review – A High Quality Workforce ....................... 52
1.3.6: Restrictions to the Number of Hours of Training - The Temple Report .... 58
1.3.7: The NHS Future Forum .................................................................. 66
1.3.8: The Shape of Training Report ........................................................ 68

1.4: Focusing on Quality in Education ....................................................... 72
1.4.1: Assessment of Quality .................................................................... 73
1.4.2: Innovation to Improve Quality ......................................................... 76

1.5: Hypotheses .......................................................................................... 83

Chapter 2: Global review of surgical curricula ............................................ 84
2.1: Introduction .......................................................................................... 85
2.2: Methods .............................................................................................. 85
2.3: Results .................................................................................................. 87
2.3.1: Australian Surgical Curriculum ......................................................... 88
2.3.2: Canadian Surgical Curriculum - CanMEDS .................................... 90
2.3.3: Hong Kong Surgical Curriculum ...................................................... 92
2.3.4: UK Surgical Curriculum - ISCP ...................................................... 93
2.3.5: USA Surgical Curriculum - SCORE ............................................... 95
2.3.6: Comparison with GMC Standards .................................................. 99
2.4: Discussion ........................................................................................... 103
2.5: Conclusion .......................................................................................... 114

Chapter 3: A Comparative Study of Contrasting Surgical Residency Programs ........................................ 116
3.1: Introduction .......................................................................................... 117
3.2: Methods .............................................................................................. 117
7.2.5: Training to Competence ................................................................................................. 217
7.2.6: Randomisation .................................................................................................................. 218
7.2.7: Control .............................................................................................................................. 218
7.2.8: Intervention: Video-Based Coaching .................................................................................. 218
7.2.9: Porcine laparoscopic cholecystectomy .......................................................................... 220
7.3: Outcome measures ............................................................................................................. 220
7.4: Statistical analysis ................................................................................................................ 221
7.5: Results .................................................................................................................................. 222
7.5.1: Demographics .................................................................................................................... 222
7.5.2: Baseline laparoscopic skills ............................................................................................. 222
7.5.3: Virtual Reality laparoscopic cholecystectomies ............................................................... 223
7.5.4: Porcine laparoscopic cholecystectomy ......................................................................... 227
7.6: Discussion ............................................................................................................................ 229
7.7: Conclusion ............................................................................................................................ 233

Chapter 8: An Immersive and Intensive Simulation Week “Boot Camp” to Improve the Quality of Surgical Training ........................................................................................................... 235

8.1: Development of SIMweek – An Intensive & Immersive Simulation Week “Boot Camp” ........237
8.1.1: Introduction ........................................................................................................................ 237
8.1.2: Methods ............................................................................................................................ 238
8.1.3: Results ................................................................................................................................ 242
8.1.4: Conclusion .......................................................................................................................... 256
8.2: Pilot Testing of An Intensive and Immersive Simulation Week “Boot Camp” to Improve the Quality of Surgical Training ............................................................................................. 257
8.2.1: Introduction ........................................................................................................................ 257
8.2.2: Methods ......................................................................................................................... 258
8.2.3: Statistical Analysis ........................................................................................................ 268
8.2.4: Results ........................................................................................................................... 268
8.2.5: Discussion ....................................................................................................................... 276
8.2.6: Conclusion ....................................................................................................................... 280

Chapter 9: Conclusion ........................................................................................................ 281

9.1: Aims .................................................................................................................................. 282
9.2: Hypotheses ...................................................................................................................... 283
9.3: Discussion ......................................................................................................................... 284
9.4: Limitations ........................................................................................................................ 292
9.5: Future Work ...................................................................................................................... 295
9.2: Concluding Remarks ........................................................................................................ 299

References .............................................................................................................................. 303

Appendices .............................................................................................................................. 320

Appendix A: Surgical Education Survey Questions ................................................................. 321
Appendix B: Quality of Surgical Training Semi Structured Interview Protocol ...................... 328
Appendix C: Statements not reaching pre-determined criteria for acceptance ......................... 332
Appendix D: Sample SIMweek OSCE Score Sheet ................................................................ 334
Appendix E: SIMweek Confidence Questionnaire ................................................................... 335
List of Tables and Figures

Figure 1a: Steps to Improve Surgical Training Quality ................................................................. 29
Figure 1b: Timeline of Government Reports Affecting Surgical Training ...................................... 57
Figure 1c: Postgraduate Competitive Entry Points Into Surgical Training and Mandatory
Examinations ................................................................................................................................. 82
Figure 2a: Standard Surgical Training Pathway ............................................................................. 87
Table 2a: Comparison of Surgical Curricula Against GMC Standards ........................................... 98
Table 2b: The ideal surgical curriculum constructed from the best components of the curricula
examined ........................................................................................................................................ 108
Table 3a: Demographic information, operative case volume and attending supervision level of
London and UPenn residents .......................................................................................................... 120
Table 3b: Comparison of activities in a typical day or night on-call as recalled by London and
UPenn residents reported as a percentage of total shift ............................................................... 122
Table 3c: Clinical activities as reported by London and UPenn trainees ....................................... 124
Table 3d: Access and exposure to educational activities as reported by London and UPenn
trainees ............................................................................................................................................. 125
Table 3e: Feedback from attendings as reported by trainees on a 5-point Likert scale (1-Never,
3-Sometimes, 5-Always) ................................................................................................................ 126
Table 4a: Top 10 indicators of quality in training ............................................................................. 141
Table 4b: Top 10 Methods of Measuring Quality in training ............................................................ 145
Figure 5a: International Representation of Surgical Education Experts ........................................... 165
Table 5a: Statements reaching pre-defined consensus criteria for acceptance ......................... 166

Figure 5b: Surgical training Quality Assessment Tool (S-QAT) .................................................. 168

Figure 6a: Written Feedback Questionnaire .............................................................................. 185

Table 6a: Written Feedback from the Interactive Workshop ...................................................... 187

Table 6b: Written Feedback from the Interactive Workshop – Free text responses ............... 188

Table 6c: High Quality Training As Rated by S-QAT responses .............................................. 198

Table 6d: Potential Areas for Training Quality Improvement As Rated by S-QAT responses .... 199

Table 6e: High Variability in Training Quality As Rated by S-QAT responses ......................... 200

Table 6f: Low Variability in Training Quality As Rated by S-QAT responses ......................... 201

Figure 7a: The 8 questions coaching framework for surgery developed with the GROW model(Alexander and Renshaw 2005; Landsberg 1997; Passmore 2010; Whitmore 1992). ................................................................. 214

Figure 7b: Flow of subjects through the study ......................................................................... 215

Table 7a: Virtual Reality Laparoscopic Cholecystectomy Global Rating Scores (a) GOALS, (b) OSATS, (c) OPRS ................................................................. 224

Figure 7c: The quality of virtual reality laparoscopic cholecystectomy performance as assessed by OPRS ................................................................. 225

Table 7b: Virtual Reality Laparoscopic Cholecystectomy Simulator metrics (a) Total Time Taken, (b) Total Number of Movements, (c) Total Path Length ............................... 226

Figure 7d: The quality of porcine laparoscopic cholecystectomy as assessed by OPRS ........ 228

Figure 8a: Flow Diagram of Search Strategy .............................................................................. 240

Table 8a: Characteristics of studies in literature review ............................................................... 241
Table 8b: Common Mock Pages and relevance to Foundation Programme competencies...... 247

Figure 8b: (i) Simulated Operating Suite (ii) Simulated Ward .............................................. 250

Table 8c: Sample Timetable ........................................................................................................ 252

Table 8d: Costs for the SIMweek program using a typical Academic Surgical Residency program

with 40 interns covering all surgical specialties as an examplea ........................................ 255

Table 8e: SIMweek Timetable........................................................................................................ 260

Figure 8c: Example Mock Bleep Scenario................................................................................. 263

Figure 8d: Study Design................................................................................................................ 265

Table 8f: Primary Outcomes........................................................................................................ 270

Figure 8e: (i) Clinical Performance - Clinical OSCEs................................................................. 271

Figure 8e: (ii) Clinical Performance - Surgical Technical Skills .............................................. 272

Table 8g: Secondary Outcomes................................................................................................... 274

Figure 8f: Self-reported Confidence............................................................................................ 275
Contributions

All research described in this thesis represents work carried out by myself. All studies were overseen by my supervisors Professor Ara Darzi and Mr Rajesh Aggarwal who both made significant contributions to all stages of the studies reported. In addition, the work was often supported by a number of individual collaborators who I would like to thank and acknowledge below.

Chapter 1: Introduction concept by Pritam Singh (PS). Written by PS.

Chapter 2: Study concept and design by PS. Data extracted by PS. Data analysis by PS. Study written by PS.

Chapter 3: Study concept and design by PS. Data collection by PS, Daniel Hashimoto (DH) and Noel Williams (NW). Data analysis by PS and DH. Study written by PS.

Chapter 4: Study concept by PS. Design of the study by PS with contributions from Sonal Arora (SA). Data collection by PS. Data analysis by PS with contributions from Philip Pucher (PP) and Alexandra Duisberg (ALD). Study written by PS.
Chapter 5: Study concept by PS. Design of the study by PS with contributions from Boris Zevin (BZ) and Teodor Grantcharov (TG). Data collection by PS. Data analysis by PS with contributions from BZ. Study written by PS.

Chapter 6: Study concept and design by PS. Data collection by PS and Timothy Allen-Mersh (TAM). Data analysis by PS. Study written by PS.

Chapter 7: Study concept and design by PS. Data collection by PS and Muaaz Tahir (MT). Data analysis by PS, with contributions from MT and PP. Study written by PS with contributions from MT (methods and results).

Chapter 8:

8.1: Study concept and design by PS. Data collection and analysis by PS with contributions from PP. Study written by PS.

8.2: Study concept and design by PS. Data collection overseen by PS, but with contributions from DH, PP, Laura Beyer-Berjot (LB), Rasiah Bharathan (RB), Katherine Middleton (KM) and Joanne Jones (JJ). Data analysis by PS with contributions from PP. Study written by PS.

Chapter 9: Conclusion concept by PS. Written by PS.
Research Outputs

The following presentations and publications relate to work from this thesis to date:

Publications

Chapter 1

1. Singh P, Darzi A. Surgical Training. Themes from the introduction have been published in this invited commentary on surgical training.


Chapter 2


Chapter 3


Chapter 4


Chapter 5


Chapter 7


*Annals of Surgery* 2014 Sep 1. [Epub ahead of print].

Chapter 8


Presentations

Chapter 3

1. Singh P, Aggarwal R, Hashimoto DA, Williams NN, Darzi A. A Comparison of Surgical Training Quality Between the UK and the USA.

Chapter 4


Chapter 5

   Poster presentation at the International Surgical Congress of the Association of Surgeons of Great Britain and Ireland, Harrogate, UK, April-May 2014.


**Chapter 6**


Chapter 7


Oral presentation at the 7th Annual American College of Surgeons Accredited Education Institutes Consortium Meeting, Chicago, Illinois, USA, March 2013.

Chapter 8


Accepted for poster presentation at the International Conference on Residency Education, Toronto, Canada, October 2014.


Poster presentation at the International Surgical Congress of the Association of Surgeons of Great Britain and Ireland, Harrogate, UK, April-May 2014.


Poster presentation at the Association for Surgical Education, Surgical Education Week, Florida, Orlando, USA, April 2013.
Chapter 1: Introduction

Themes from the introduction have been published in this invited commentary on surgical training:

1.1: Aims of the Thesis

- To identify variations in the quality of global surgical training programs

- To develop an assessment tool to objectively evaluate quality in surgical training programs.

- To develop and pilot interventions that will enhance the quality of surgical training programs.
1.2: Overview & Rationale

Quality assessment has become a major principle of healthcare. Clinical and patient related outcomes are routinely collected and analysed. However, the assessment of surgical training quality is not as well established. This thesis will investigate the hypothesis that quality of surgical training cannot be measured by quantity alone and subsequently develop an assessment tool that can robustly assess the quality of surgical training. The later chapters will focus on the development of interventions to improve the quality of surgical training to make it more efficient.

The next section of this introduction chapter will examine the history of surgical training using the UK training system as a focus point. Since the turn of the millennium there has been increasing political interest in the structure and organisation of postgraduate medical education. This chapter examines the effects of the government consultations on training and summarises the increasing focus on quality of training.

Following on from the focus on the UK in chapter 1, chapter 2 of this thesis will assess the current state of surgical training globally with a review of the organisational structure of current international surgical training programs. In order to identify indicators of high quality surgical training, it is first important to identify the variations in surgical training programs. An online search will be performed for nationally recognised general surgery curricula followed by a qualitative evaluation of each curriculum. This will be followed by assessment of each curriculum against the ‘Standards for curricula and assessment systems’ produced by the
General Medical Council. This study will identify strengths and potential areas for development within existing general surgery curricula. In addition it will provide information about the variability of the currently employed general surgical training programs globally.

Chapter 3 will focus in on the general surgical training programs of two of the nations examined in Chapter 2. A comparative study of the quality of surgical training as reported by trainees in the UK and the USA is described. This is a collaborative study with the University of Pennsylvania. In the UK, the Joint Committee on Surgical Training (JCST) has developed quality indicators for surgical training. In the USA, similar indicators can be inferred from the American Board of Surgery (ABS) and Accreditation Council for Graduate Medical Education (ACGME) Surgical Residency program requirements. A questionnaire will be developed based on the themes from the JCST quality indicators and distributed to trainees in both countries.

In order to define the performance metrics of a high quality surgical training program there is first a need to identify the variables that need to be assessed. In Chapter 4 of this thesis, a semi-structured interview study sampling both trainees and trainers will be conducted to identify and define markers of the quality of surgical training and discuss the metrics for measuring quality in surgical training. A purposive sampling method will be employed to ensure representation of all levels of seniority and experience in order to fully explore the trainee-trainer relationship. The final objective of this study will be to develop concepts for interventions that can be used to improve the quality of surgical training.
Building on themes gathered from Chapter 4, in Chapter 5 a Delphi consensus methodology will be employed to develop international expert consensus on the indicators of high quality surgical training. An international panel of experts in surgical education will be invited to participate from a number of countries. This will include trainers and representatives of national trainee organisations. The results of Chapter 5 will facilitate the development of an assessment tool in Chapter 6 that will seek to objectively rate the quality of a surgical training program. The prototype tool will be refined during international focus groups or workshops in order to gauge opinion on usability and how best to pilot, evaluate and finally implement the tool.

Identification of indicators of high quality surgical training in the early parts of this thesis will subsequently allow focused training interventions to be developed that can maximise the quality of surgical training. In Chapter 7 and Chapter 8, training interventions developed to target improvements in surgical training quality will be studied in more detail. Chapter 7 will assess a training intervention at the individual trainee level, while Chapter 8 will be used to develop an intervention at the institutional level. These intervention studies will be conducted in a hybrid simulation and clinical setting and will be able to call upon the institutions’ experience with the use of simulation in surgery.

This thesis will facilitate the development of robust evaluation of the quality of General Surgery training programs. This will enable evidence-based recommendations to be made to improve the quality of the General Surgery training program. Reliable assessment of quality will permit
evaluation of new innovative interventions in surgical training programs both nationally and globally. Furthermore, longitudinal assessment of the quality of surgical training may in itself have a positive effect in driving up the standards of surgical education.
Figure 1a: Steps to Improve Surgical Training Quality

To improve the quality of surgical training, there is a need to develop a method to measure training quality.
1.3: The Evolution of Surgical Training

For over a century surgical trainees have adopted the apprenticeship model propagated by William Halsted from the Johns Hopkins University School of Medicine. His ‘see one, do one, teach one’ method of teaching has been recited by surgeons for generations. However, over the past decade attitudes towards the subject of surgical training have changed. The pressures of time, innovation and introduction of new technology, and increasing awareness of patient safety requirements have been drivers of this change (Lewis and Klingensmith 2012). The cumulative effects of reduced hours, reduced operative volume and surgical subspecialisation has forced a paradigm shift from time and volume towards competency-based training. Traditionally a well-filled operative logbook acted as an indicator of competence. Training improvements focused on increasing quantity through time in training or caseload. The move towards competency-based assessment has required a similar but challenging shift in focus from the number of hours of training to the quality of those training hours.

Volume of training through time in training or through caseload has been employed as a surrogate indicator of quality. In support of this there exists robust evidence that improved patient outcomes can be correlated with increased operative volume (Birkmeyer et al. 2003). The volume-outcome relationship of surgeons (Birkmeyer et al. 2003) and hospitals (Birkmeyer et al. 2002), has transformed the geographical distribution of surgical procedures. As a result, the preceding decade has seen a dramatic increase in subspecialisation both of surgeons, but more importantly for training, of hospitals. For example the impact of centralisation of vascular aortic procedures (Forbes 2011) and upper gastrointestinal cancer resections (Siriwardena 2007)
will indirectly have effects on the exposure to these specialties for trainees not working in these specialised units. Consequently structured and targeted training is now crucial, as time alone may no longer ensure sufficient exposure to the breadth of training required to produce a competent surgeon.

In fact, what is required from a competent surgeon is an interesting question. Traditionally general surgical and specialist surgical training has been completed within the higher surgical training programme. However, centralisation, subspecialisation and the introduction of new technologies have led to the increased uptake of fellowships. In the USA more than 80% of graduating general surgical residents are going into fellowships (Lewis and Klingensmith 2012). A recent national survey of UK surgical trainees found that over 75% had already undertaken or were intending to undertake a fellowship at the end of their training (Fitzgerald et al. 2013).

Advances in minimally invasive surgery have certainly created challenges for training. They have also been responsible for an exemplary structured, competency-based training programme. In 2006, NICE recommended laparoscopic resection as an alternative to open resection for colorectal cancers. However, due to a lack of adequately trained colorectal surgeons to perform the laparoscopic resections, the guidance could not be followed. The LAPCO national training programme (Coleman) was developed to address this shortage of surgeons competent to perform laparoscopic colorectal surgery. It combined pre-clinical and clinical training with end of programme assessment to confirm competency. This innovative example of training surgeons facilitated the safe and successful introduction of a new surgical technique. In the USA
The Fellowship Council has been created to develop and maintain high quality surgical fellowships (FC). The body was originally created to oversee the introduction of minimally invasive surgery fellowships, however it now covers multiple specialties including bariatric, thoracic and flexible endoscopic fellowships. Program accreditation and reaccreditation requires a combination of online submission of information and site visits by independent reviewers.

Training can be unintentionally hampered by politically driven financial incentives for service delivery. Financial incentives for maintaining a high volume, high standard of service delivery have been associated with a reduction in mortality (Sutton et al. 2012). The long term benefits of these financial incentives remains questionable (Kristensen et al. 2014). However, in the short term this combination of financial and clinical gains can create a conflict between service delivery and training provision. In the UK, the public inquiry following failures in the Mid Staffordshire NHS Foundation Trust found that ‘the Trust was operating in an environment in which its leadership was expected to focus on financial issues….. Sadly, it paid insufficient attention to the risks in relation to the quality of service delivery this entailed’ (Francis 2013). The long-term view should be encouraged, as there is evidence that patient outcomes are related to the quality of training received. A retrospective analysis of almost 5 million cases involving over 4000 clinicians from 107 US residency programs, demonstrated that programs could be ranked according to the risk-adjusted complications of the patients treated by the graduates of those programs (Asch et al. 2009). Patients treated by a graduate from the bottom quintile residency programs experienced an approximately 30% higher complication rate than
those treated by graduates of a top quintile program. This provides empirical support that the quality of training received can be associated with patient outcomes. This concept was also supported in the recommendations from the Francis inquiry which highlighted that ‘medical education and training systems provide an opportunity for enhancing patient safety’ (Francis 2013). Therefore improvements in surgical training could potentially lead to a higher quality patient service. This seems intuitive and indeed in the UK the previous two decades have seen a great number of major efforts to reform postgraduate medical education which are described below. Analysis of their implementation and in particular some of the challenges posed after their implementation provides support for the necessity to measure training quality. This is expanded upon in section 1.4.

1.3.1: Calman Reforms

In the UK, since the implementation of the Calman reforms in the 1990’s, there has been great interest in the quality of medical training. The Chief Medical Officer Sir Kenneth Calman proposed the reforms with the publication of ‘Hospital Doctors – Training for the Future’ in 1993. The main objective of the reforms was to improve training for hospital specialists with more structured and intensive training. In 1996 the Specialist Registrar training post was introduced to replace the previous Registrar and Senior Registrar training posts. The abolishment of the Senior Registrar grade was a response to the large numbers of trained doctors awaiting a post, which focused on experience rather than targeted training. The Specialist Registrar post was intended to reduce the time spent in training from an average of 12 to a minimum of 7 years. Importantly a definitive end point of training was created with the
proposed Certificate of Completion of Specialist Training (CCST), which was to be awarded by the General Medical Council (GMC). The key principles of the new Specialist Registrar post were:

- Explicit educational and service entry requirements into the training program
- Competitive entry
- Shorter duration
- Intensive, managed training
- Progression dependant on educational attainment
- Graded service responsibility as the trainee progresses (K. Calman 1995)

Since the Calman reforms were focused on the Senior Registrar grade, the earlier years of training including the Pre-Registration House Officer (PRHO) and Senior House Officer (SHO) grades remained unaffected.

### 1.3.2: The ‘Lost Tribe’

Following the implementation of the Calman reforms, in 2002 the Chief Medical Officer Sir Liam Donaldson published *Unfinished Business – Proposals for reform of the Senior House Officer grade* (L. Donaldson 2002). This was a response to concerns that the Calman reforms had inadvertently created problems further down the training pathway at the Senior House Officer (SHO) grade. Nineteen proposals were described and set out for consultation. The Calman reforms (K. C. Calman et al. 1999) had focused on senior specialist trainees traditionally known as registrars. The pre-registration house officer (PRHO) year had been targeted with the General Medical Council’s *The New Doctor Recommendations on General Clinical Training* (GMC
The subsequent early years of training during the Senior House Officer (SHO) grade had not been reformed. The Senior House Officer position was usually sought by doctors for their second postgraduate year of service upon successful completion of the PRHO grade. This was usually the point in the training scheme at which doctors would begin to specialise in a field such as medicine or surgery as PRHO appointments had strict requirements for generic training in both medicine and surgery in order for full registration to be awarded. *Unfinished Business* (L Donaldson 2002) highlighted the need for greater career structure and the establishment of explicitly defined endpoints for SHOs. It was the absence of the latter that led to SHOs becoming referred to as the ‘lost tribe’. These were trainees who through failure to secure competitive entry onto a Specialist Registrar programme, continued to work for indefinite periods, sometimes even a decade, as SHOs. This was essentially leading to a repetition of training, which would certainly provide trainees with extra experience but would rarely count towards the higher specialist training years. It was noted that short-term jobs independent of any training programme comprised half of all SHO appointments. Furthermore, where training jobs had been arranged as a ‘rotation’ these were still unlikely to be delivered as a managed programme. There was a lack of standardisation in selection procedures for SHO appointments with a failure to build these around clearly defined core competencies. In addition the report highlighted the lack of supervision and inadequacy of appraisal procedures. This was likely compounded by the absence of clear competencies and assessments. An SHO’s competence to progress to the coveted specialist registrar grade was often judged on their success at securing further SHO posts and completion of the relevant Royal College examinations. In identifying the increasing workload placed on SHOs, the report acknowledged the magnitude of service
delivered by SHOs in the NHS. The report highlighted that fifty percent of NHS medical trainees were SHOs and that a third of these were overseas graduates.

Drawing from the experience of the recently proposed Calman reforms, the recommendations for the reform of the SHO grade centred on five principles which are subsequently discussed:

1. Programme based
2. Time limited
3. Broad based at the start
4. Flexible to permit trainees to move into, out of and between training schemes
5. Tailored to meet individual needs.

A 2-year Foundation Programme was proposed in between graduation from medical school and the commencement of basic specialist training programmes. The pre-registration house officer year would be included in the Foundation Programme. The objective was to facilitate acquisition of generic skills that would be required in all specialties and to provide 2 years of work experience to inform future career choices. The basic specialist training programmes would follow the five principles of reform and hence be time limited. The programmes would be designed with the impending full implementation of the European Working Time Directive requirements in mind. With the resultant need for significant changes in working patterns, the need for a focus on enhancing the quality of training was noted:
'The way forward is to acknowledge that trainees will be available for shorter periods and will be engaged in new working patterns and then to build new approaches to training. Such approaches should increase its quality, the opportunities to learn and the ways in which skills and training are acquired.'

Although the report’s initial objectives were to focus on the early years of training, the considerable reforms proposed for the SHO grade led to discussion about the links with higher specialist training. The report introduced the early concept of ‘run-through’ training that would feature more prominently in the following years. The concept was of a single training grade in which progression would be possible without the need for application to another programme as long as trainees met performance and assessment competencies. The long training programme to reach the consultant grade when compared with European models was mentioned. The potential to shorten the length of training was muted. This would produce a more generalist consultant with the option to further specialise. There was acknowledgement that run-through training was not ready for immediate implementation. However, the proposal of the concept in itself appeared contradictory to the principle of increased flexibility. This would become more apparent in the years following the report’s adoption and implementation.

The overall control of the proposed training programmes was to be moved away from individual hospitals and instead postgraduate medical deans would be given control of appointing, assessing and providing the training for the trainees in the new programmes. As
part of the reforms, in-training assessment was to follow the principles of competency-based assessment. A formalised assessment and appraisal system would therefore provide evidence to support successful graduation from the training programme. Importantly, postgraduate deans would also be responsible for training and providing support to the trainers in an attempt to provide quality assurance for the provision of training. Postgraduate deans would be responsible for ensuring selection procedures comply with nationally agreed standards. It was recommended that agreed standards for both selection and assessment should be published and made available for trainees.

1.3.3: Modernising Medical Careers

Sir Liam Donaldson’s proposed reforms of the SHO grade were largely well received. As a result, a policy statement on Modernising Medical Careers (MMC)(DoH 2003) was published in February 2003 outlining major reforms to postgraduate medical training. Other drivers for the reforms included the impending implementation of the European Working Time Directive and the commitment in the NHS Plan (2000) towards a more consultant delivered service.

In October 2003 a strategy group was assembled to commence implementation of the Modernising Medical Careers (MMC) reforms which culminated in the publication of MMC: The next steps(DoH 2004) in April 2004. The group was led by Sir Liam Donaldson and included representation from the four UK Health Departments, the General Medical Council, Joint Committee for the Postgraduate Training of General Practitioners, the Medical Royal colleges, the Conference of Postgraduate Medical Education Deans of the UK. The Next Steps report was
intended to further provide details regarding the principles proposed in the original MMC Policy statement. The intention was for the programme to provide a broader perspective of the potential career opportunities available to junior doctors. The original principles of the Foundation Programmes were reasserted:

- Trainee centred
- Competency assessed
- Service based
- Quality assured
- Flexible
- Coached
- Structured and streamlined

The MMC reforms embraced the concept of the Foundation Programme; a two-year internship commenced immediately post graduation. Foundation Year 1 would be a pre-registration year that would necessitate demonstration of training outcomes in order for the trainee to successfully apply for full registration at the end of the year. The programme was also to incorporate a competency based assessment framework capable of providing evidence of improvement in competencies from the first to the second year.

Plans were outlined to Pilot the Foundation Programme prior to its implementation in August 2005. Given the proposed timescales, it was an essential requirement that the programmes were compliant with the European Working Time Directive. The value and necessity for piloting
such major reforms to training were clear. However, the *next steps report* was published in April 2004 and the junior doctors in question rotate once a year in August. Therefore this left little time to develop and perform the pilots and almost no time for evaluation before the full implementation in August 2005. In fact the Foundation Programme curriculum was published in June 2005. Legal responsibility for approval and quality assurance of the Foundation Programmes was to be given to the General Medical Council (GMC) and the Postgraduate Medical Education and Training Board (PMETB).

Selection and progression criteria for Foundation programmes would be transparent with competitive selection nationally for all posts. Progression would be based on the assessments of the GMC. In perhaps the most fundamental of the reforms outlined, following completion of the Foundation Programme, rather than looking for SHO appointments, trainees would be expected to enter specialist training programmes leading to a Certificate of Completion of Training (CCT). The principle of ‘run through’ specialist training had been formally adopted.

*Unfinished Business* introduced the concept of run-through training with the acknowledgement that it was not ready for immediate implementation. However, following the consultation on the proposed reforms, the *Modernising Medical Careers* team were much more enthusiastic about the run-through approach and felt it was ‘not only desirable but also achievable.’ This would result in a single competitive entry point into specialist training at the end of the Foundation Programme. If all competencies were met and progression was satisfactory then no further competitive selection would be required for completion of the training programme. The
single competitive selection point would need to be before the end of the Foundation Programme in order to ensure direct entry into the specialist-training programme. For this reason prior experience in a chosen specialty was not a criteria for selection as some Foundation trainees may not yet have rotated into that specialty. An alternative interpretation was that experience, ability and commitment to a chosen specialty would no longer provide a candidate with any competitive advantage when compared with an inexperienced trainee with no demonstrable aptitude in a given specialty, in fact one in which they may never have received exposure to. For general surgery this was for appointment to an 8-year specialist-training programme leading towards CCT. The selection procedure for this run-through training programme was to become the most controversial and least acceptable of the MMC reforms (Brown et al. 2007). Furthermore, the route back into training for those who were unsuccessful at the competitive entry point was not clear.

Progression through the specialist-training programme was again intended to be based on competency-based assessments. Upon satisfactory completion of the programme and the appropriate evidence of competence, the trainee would be awarded a CCT and be eligible for a consultant post. However, the report was sympathetic to the concept that in this new specialist-training programme, further subspecialist training may be required following the achievement of CCT.

The appendices of the report indicate a greater interest in quality assurance of training. It was formally noted that high quality training would be essential to maintain a high quality of patient
care. All programmes would need to have a clearly defined curriculum complete with a systematic series of assessments of both the formative and summative nature. The standards required for competency needed to be transparent and in the public domain. One of the principles outlined was that the programmes were required to have explicit quality assurance processes. While this was a commendable principle, little detail or guidance was provided on how this would be achieved.

1.3.4: Implementation of MMC reforms

A national standardised selection process was developed for the appointment of trainees to the specialist training programmes: the Medical Training Application Service (MTAS). The Department of Health proposed plans for this in 2006 and it was to be active from January to August 2007. A standardised application form and shortlisting criteria were developed. The applications would be co-ordinated in an on-line system with interviews after shortlisting. A further round of recruitment would be used to fill any vacancies at the end of the first round.

The pilot Foundation Programmes were commenced in 2004 with full implementation nationally from August 2005. This left very little time for any meaningful analysis and feedback from the pilots to inform the full implementation. Furthermore, given the first cohort of Foundation Programme trainees would complete in 2007, the decision to implement the sweeping reforms to the specialist training programs and commence these at the same time in August 2007 seemed ambitious from the outset. The fact that no formal pilot was planned
before implementation of the reformed specialist training programmes suggested a lack of appreciation of the scale of these reforms.

What followed was one of the most turbulent periods in the history of medical education in the UK with high profile criticisms of the way the reforms were implemented (Brown et al. 2007; Nicholl 2007). The initial MTAS review group led by Professor Neil Douglas described it as

‘The biggest crisis within the medical profession in a generation (DoH 2007).’

One of the defining moments of the unrest caused by the implementation of the MMC reforms came in March 2007 when the interview panel selected to conduct the general surgery specialist training programme interviews in the West Midlands Deanery objected. The refusal to conduct the interviews was on the grounds of serious concerns with the short-listing procedures. This was shortly followed by demonstrations in Glasgow and London by thousands of junior doctors (BBC 2007). There were widespread calls for an independent review, which was eventually established by the Secretary of State for Health in April 2007 and performed under the leadership of Professor Sir John Tooke.

1.3.5: Aspiring to Excellence

The well-documented criticisms (Brown et al. 2007; Nicholl 2007) and failures during the MMC implementation led to an in depth enquiry led by Sir John Tooke into the framework and processes underlying MMC. His interim report ‘Aspiring to Excellence’ was published in October
2007 and following a consultation the final report was published in January 2008 (Tooke 2008).

As indicated by the title, the report starts with a clear message that reforms to ensure training was ‘good enough’ needed to be superseded by aspirations for training to be excellent.

The independent inquiry performed forensic analysis of all formal documentation of meetings and relevant publications relating to MMC. Oral and written evidence was solicited and received from 130 organisations involved with MMC. Key features of the inquiry were eight trainee workshops arranged across the UK to sample views from those most affected by the reforms. Trainees were asked about:

1. Benefits and disadvantages of run-through training and what solutions they would suggest?
2. Whether or not they felt Foundation Year 2
   a. Added to Foundation Year 1?
   b. Allowed them to experience their chosen specialty of interest?
   c. Was valuable?
3. Methods to make non-training grades more attractive as a career option?
4. How to address the mismatch between demand for higher specialist training and the workforce and service requirements for specialists?

The strength of the Tooke report lay in its depth and comprehensive nature. The background and context in which the MMC reforms had been developed were examined and summarised.
comprehensively. A number of key developments were identified that acting synergistically created an environment where postgraduate medical education and training was sinking in the order of healthcare priorities. The first of these was the significant structural reorganisations of the administration of the NHS in England, which favoured a market and more commercial discipline. This led to a management culture with tightly controlled performance measures linked to national targets. The problem was that these targets focused on financial performance, clinical standards and service delivery while not directly covering training and educational targets. An example was the ‘payment by results’ concept whereby tariffs were created for units of clinical service. It was logical that a well trained medical workforce would improve clinical and service delivery standards, however this would be a long term view that would be of little interest to a management team under pressure to meet annual national financial and service targets.

The second set of developments identified in the Tooke report was the progressive regionalisation of responsibility for the management and funding of postgraduate medical education. In 2006 ten Strategic Health Authorities (SHAs) were responsible for the delivery of postgraduate medical education alongside their service delivery targets. The MMC reforms were developed with a national standardised system in mind but needed to be delivered within this regionalised management structure. In addition, financial constraints had led SHAs to reduce their education and training budgets to maintain their overall financial health. This was made possible because the new administrative structures no longer ring-fenced the Multi-Professional Education and Training Levy (MPET). Concerns were acknowledged over the
vulnerability of funding for medical training and in fact it was considered to be of great risk in

Whilst postgraduate medical education funding was falling down the priority list and losing its
protection, the 2006 Review of UK Health Research Funding proposed ring-fenced government
funding for health research. Furthermore, allocation of funding was facilitated by the well-
established methodology of assessing research output in terms of publications and citations.
Similarly the introduction of explicit targets for financial, clinical and service delivery had
objectively measurable outcomes to justify allocation of resources to them. Postgraduate
medical education had neither the ring-fenced budgets afforded to research funding, and
furthermore the quality of postgraduate medical education had no established methodology to
quantify output from hospitals or universities.

In reviewing the policy objectives for the MMC reforms, there were some positives to be taken
from the introduction of the Foundation Programme. A formalised national curriculum with a
framework for competency-based assessments was an achievement for a previously informally
structured year. The selection process was however felt to give too little weight to performance
at undergraduate examinations and assessments. The adequacy of integration between
Foundation Year 1 and undergraduate teaching programmes were another are for
improvement identified during the consultation. Perhaps unsurprisingly, a formal evaluation of
the four pilot Foundation Programmes in Kent, Surrey and Sussex found that the majority of
trainees felt the implementation of the curriculum was rushed and they echoed the e-
consultations findings that integration with undergraduate training could be improved (Dewhurst et al. 2006).

The inquiry’s findings were summarised into eight key areas where action was required. Recommendations were made in each of these areas and are discussed below:

1. The policy objectives for MMC needed to be clearly defined and agreed by the key stakeholders.

The original principles of Unfinished Business had been adapted to an extent that some founding principles were lost or contradicted. For example, the principle of increasing flexibility gave way to plans to implement ‘run-through’ training programmes. Furthermore, the fate of trainees assigned fixed term training appointments, was poorly defined. This led to concerns that these fixed term appointments may well evolve into another ‘lost tribe’.

Recommendations were made to redefine the principles that would underpin medical education and training, while basing these on the original ideas from Unfinished Business. Policy development should be based on evidence where available. The Department of Health should engage with and formally consult the medical profession on all significant changes to policy that affects the medical workforce, medical education and service delivery. In order to implement these recommendations creating a new organisation; NHS Medical Education England would be valuable.
2. **The role of the skills and duties a doctor needed at the different grades needed more clarification.**

The consultation revealed concerns that doctors were taking longer to gain increased responsibility under supervision. Concerns were raised about the capabilities of a trainee once they achieve CCT. Some training opportunities were being lost to other healthcare professionals within the multi-disciplinary team. This was often due to the service benefits of such substitutions.

3. **The governance of and accountability structures of the Department of Health’s policy development needed to be strengthened.**

Governance was compromised by ambiguity in the accountability structure, in particular with respect to the national recruitment process. The Medical Training Application Service (MTAS) was implemented with a huge number of changes introduced simultaneously and without sufficient piloting. Recommendations included clearly separating policy development from policy implementation and a need to employ professional project management principles when implementing policy. The consultation revealed a strong desire for the Chief Medical Officers to be accountable for medical education and this was formally incorporated in to the recommendations.

4. **Workforce planning requires optimisation.**

The complex nature of workforce planning was acknowledged, as was the need for consistent workforce policies. The ‘run-through’ training principle failed both to provide the broad based
early training proposed under the original MMC plans and its rigidity hindered future workforce planning capacity to respond to evolving technological advances and service requirements. Recommendations were to nationally centralise higher specialist training commissioning and award Trusts on a competitive basis taking into account service, education and research records.

5. There needs to be an increase engagement from the medical profession for training policy development.

The consultation process revealed concerns that doctors were becoming ‘deprofessionalised’ and they needed to be more involved in developing and implementing training related policies. There was acknowledgment that the royal colleges and the profession had not always provided the leadership required to deliver a coherent message to inform training policy development. It was recommended that opportunities should be made available to medical trainees to acquire management skills to facilitate the increasing need for collaboration between clinicians and managers. Trainees also needed to be represented in Trust management structures. The proposed NHS:MEE body was recommended to coordinate the production of advice on postgraduate medical training.

6. Optimising the commissioning of postgraduate medical education.

The restrictions to the training budgets imposed by the SHAs in England in 2006/7 demonstrated the vulnerability of non ring-fenced training budgets in the presence of financial pressure in the NHS. Better incentives were required to improve medical education. It was
recommended that medical education should be based on evidence and thorough critical appraisal. This could be encouraged through better collaboration between the NHS service sector and higher education institutions. The report noted evidence that there was positive correlation between Trusts that invest in education and research and Healthcare Commission ratings.

The recommendations introduced the concept of medical education funding reflecting not only the training requirements but in addition the service and academic contributions. One of the specific recommendations was to incorporate postgraduate medical training and education into the Healthcare Commission’s performance reports in order to provide an incentive for trusts to invest in education. This recommendation was agreed to or strongly agreed by over 88% of respondents during the consultation process. Furthermore, it was recommended that Trust Medical Directors should be made responsible for the local delivery and regular reporting of education and training. Despite all this enthusiasm, the closing comment of this section made clear the challenges facing raising the profile of medical education:

“The adoption of performance measures to incentivise Trusts to prioritise PGMET remains unresolved and needs to be addressed.”

Commissioning, incentivising and reporting of medical education and training required the development of adequate performance measures.
7. Medical education regulation should be streamlined into one body

Undergraduate medical education, Foundation Year 1, Continuing Professional Development (CPD) and revalidation is the responsibility of the General Medical Council (GMC) while the Postgraduate Medical Education and Training Board (PMETB) oversaw postgraduate medical education beyond Foundation Year 2. The recommendation was to merge the PMETB within the regulatory structure of the GMC.

8. Future structure of postgraduate medical education

For the Foundation Programme, the report recommended uncoupling Foundation Years 1 and 2, and greater integration between Foundation Year 1 and the final year of the undergraduate curriculum. In particular, the panel felt that the supervised element of FY1 experience should be drawn into the final year of medical school training. The panel further recommended the introduction of core training to replace the second year of Foundation training and the first 2 years of the ‘run-through’ training, ST1 and ST2. The proposal was for a very limited number of four themes for core training such as medicine, surgery, community and acute common stem to facilitate broad based early years training.

The recommendations concerning selection into higher specialist training took into account concerns that the MTAS selection system favoured competence over excellence with little weighting given to clinical experience, academic achievements and curriculum vitae. The development of National Assessment Centres was recommended to assess not only knowledge but also clinical skill and aptitude. The standardised national application forms would be
retained but with specialty specific questions together with the applicants curriculum vitae and assessment ranking.

The final report of the Tooke inquiry made 47 recommendations of which the Department of Health accepted 24 and addressed the remaining recommendations in the Next Stage Review. While the shortcomings of the MMC implementation were an embarrassment for the government, they highlighted a need for focused policy on medical training at time when quality was high on the political agenda in the NHS with the publication of ‘High Quality Care for All’ (DoH 2008b).

1.3.5: NHS Next Stage Review – A High Quality Workforce

In 2008 as part of the NHS Next Stage Review, the UK Department of Health published its report into the future of the NHS workforce entitled A High Quality Workforce (DoH 2008a). As the title suggests, the vision was to centre the NHS around quality. Representation was sought from professional bodies, trade unions and experts in higher education. Over 100 people took part in four working groups to develop the proposals. Their task was to consider four themes:

1. The evolving role of the healthcare professional and the implications for career structure.
2. Workforce planning
3. Education commissioning
4. Improving education to ensure high quality patient care
The core principles underlying the proposals were:

1. ‘Focused on quality
2. Patient centred
3. Clinically driven
4. Flexible
5. Valuing people
6. Promoting life-long learning’ (DoH 2008a)

The review acknowledged the changing role of the clinician from being solely a medical practitioner to the need to integrate training in leadership, management, academic research and teaching. In response to the Tooke inquiry’s request to define the role of a doctor, the report highlighted that given the on-going advances being made in healthcare systems, quality was dynamic and thus the role of the doctor would continue to evolve.

In order to provide the profession with a clear voice at national level, Medical Education England (MEE) would be established as an independent advisory non-departmental body. MEE would work with the Royal Colleges, GMC, NHS employer, universities and educational commissioners to reform the structure of postgraduate medical education. Five key structural elements were identified:
1. **Pre-registration**

Selection procedures for Foundation Programme recruitment required development to improve the validity and reliability. A key objective was to ensure all UK medical graduates would have Foundation Year 1 placements in order to achieve full GMC registration.

2. **Specialty Training**

Specialty training would commence upon completion of Foundation Training, however the details of specialty training structure were not finalised. This caution was in response to the problematic implementation of the MMC principles. MEE were tasked with finalising details in collaboration with the other training stakeholders and keeping true to the Tooke inquiry’s recommendations.

3. **Modular Credentialing**

This would allow formal credentialing of competencies throughout the training pathway. The benefits of such a system include increased flexibility to move between specialty training programmes. This would facilitate dynamic workforce planning to meet the changing service needs of society for instance in response to advances in medical technology.

4. **Partner, Leadership and Educational Roles**

Management, leadership and education skills training were to be integrated into all postgraduate training programs. In an attempt to control the quality of training, mandatory
training would be provided to educational supervisors and their performance would be reviewed.

5. **Academia and Research**

Integrated clinical and academic training pathways would be facilitated through the launch of the Academic Clinical Fellowships, and Clinical Lectureships. The schemes would follow the recommendations of the Walport report (UKCRC. and MMC) and continue to encourage clinical research careers by providing integrated and flexible training.

A central theme of this report was the desire to focus on the quality of education and training and increasing the accountability of training quality through commissioning. There was appreciation that high quality care necessitated education and training of the medical workforce to also be of high quality. Educational commissioning would permit the identification and reward of quality in education. On an individual level the report encouraged the development of empowering trainees to have greater involvement in where their training is conducted in order to optimise the quality of their education. Educational funding structure would be transformed to promote transparency and incentivise high quality education providers. This would be performed by replacing the outgoing Multi Professional Education and Training (MPET) Budget with a tariff-based system, which would be assigned to individual trainees. The focus on quality for educational commissioning would rely on clear outputs, and effective quality assurance systems. Education commissioners would be encouraged to be innovative in their approaches to incentivise high quality teaching. The report introduced the
The concept of ‘premium placement rates’ in order to encourage teaching in new settings. Research would be commissioned to identify high quality educational practice and to continue to enhance it. There would be encouragement to publish metrics on quality and expenditure for continuous professional development (CPD) of existing staff in addition to the trainees.

The report was welcomed by Professor Sir John Tooke who had led the enquiry into the concerns raised during the Modernising Medical Careers initiative. The creation of Medical Education England was something he had strongly recommended in his report Aspiring to Excellence. He also praised the efforts to both protect and incentivise investment in education through the development of a tariff-based payment system.
Figure 1b: Timeline of Government Reports Affecting Surgical Training

<table>
<thead>
<tr>
<th>Report</th>
<th>Year(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital Doctors – Training for the Future</td>
<td>1990s</td>
</tr>
<tr>
<td>Introduction of the Specialist Registrar</td>
<td>2002</td>
</tr>
<tr>
<td>Unfinished Business</td>
<td>2003</td>
</tr>
<tr>
<td>Choice and Opportunity</td>
<td>2004</td>
</tr>
<tr>
<td>Modernising Medical Careers</td>
<td>2005</td>
</tr>
<tr>
<td>Aspiring to Excellence</td>
<td>2006</td>
</tr>
<tr>
<td>NHS Next Stage Review – A High Quality Workforce</td>
<td>2007</td>
</tr>
<tr>
<td>The Temple Report</td>
<td>2008</td>
</tr>
<tr>
<td>Education and Training - next stage.</td>
<td>2009</td>
</tr>
<tr>
<td>NHS Future Forum report</td>
<td>2010</td>
</tr>
<tr>
<td>Shape of Training Review</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>2013</td>
</tr>
</tbody>
</table>
1.3.6: Restrictions to the Number of Hours of Training - The Temple Report

Since July 2011 work hours have been restricted to 80 per week in the USA (ACGME 2011a) with further restrictions of 16 hours for continuous duty for 1st year graduates and 24 hours for all others. These were more tightly restricted to 48 hours per week in the UK as a result of the European Working Time Directive (EWTD). While the EWTD was introduced in 1998, a period of adjustment was permitted for doctors in the UK and the restrictions were not implemented until the first of August 2009. However, since their implementation, both trainers and trainee associations have voiced concerns over the impact of work hour restrictions on exposure and training (Black 2009; Morris-Stiff et al. 2005; Simpson et al. 2011). The RCSEng have maintained a firm position (Black 2009) on the advantages of easing work hour restrictions for surgical trainees and there still exists intense debate over the impact of resident work hour restrictions in the USA (Drolet et al. 2012; Fletcher et al. 2005; Lewis and Klingensmith 2012; Rosenbaum and Lamas 2012).

In the UK the Secretary of State for Health requested Medical Education England to commission a review to investigate the impact of the EWTD on the quality of training. Professor Sir John Temple was chosen as the Independent Chair of the Review. He had previously been instrumental in the implementation of the Calman reforms (K. Calman 1993) and had made significant contributions to Unfinished Business (L Donaldson 2002). The Temple report published in 2010; concluded that ‘High quality training can be delivered in 48 hours’ and ‘extending the hours worked or the length of training programmes are not sustainable
solutions’ (Temple 2010). The subtitle of the Executive Summary highlighted the importance placed on the report:

‘Training is patient safety for the next 30 years.’

Given the review’s objective of assessing the impact of the EWTD on the quality of training, it was acknowledged that high quality training itself needed to be explored. Organisations and individuals were invited to submit oral and written views on high quality training. A consistent view of what high quality training should consist of included:

• Well supervised with a competency based structure
• Measurable with nationally accepted objectives
• Skilled trainers
• Avoiding excess service commitments
• Working within the trainees’ competence
• Supporting transition from supervised to independent practice
• Occur in a high quality care environment
• Patient mix should complement the curriculum

While the majority of training would need to take place in the clinical setting, some basic skills could be taught removed from patients such as in simulation scenarios. The key requirements for training placements were also discussed:

• Appropriate to the trainee’s experience
• Mandatory induction
• Employ multiple teaching methodologies
• Consultants or seniors should supervise and provide feedback
• Clear objectives
• Follow a clear curriculum
• Incorporate trainee assessment to ensure progress
• Outcomes should be monitored for continuous improvement

The review looked at information derived from four different sources; a literature review, oral and written submissions and focus groups. Forty-one organisations including professional and trainee groups participated in oral hearings. Written submissions were invited to respond to four key areas:

1. How to define high quality training?
2. What impact the EWTD has had on the quality of training?
3. How has the healthcare system responded to EWTD?
4. What can be learnt nationally or internationally about delivering training within restricted time?

Over 100 written submissions were received from a broad range of professions and organisations. A further 150 articles that were referred to in the submissions were also analysed. The analysis was completed prior to the commencement of nine focus groups to validate the findings and test the recommendations.
The review highlighted the lack of objective evidence available, which was likely to contribute to the debate on the implementation of the EWTD and its effects on training. The EWTD had in fact only been enforced (in its full 48 hour form) for 1 year. As a result where views and opinions were consistently repeated by many sources, this was used to substitute for objective evidence. Demonstrations of the effects on the outputs of training were not likely to become apparent for some years. Furthermore ‘There is a lack of nationally agreed and validated quality measures which specifically focus on the output and outcomes of training.’

In preparation for the implementation of the EWTD, medical school numbers were expanded and focus turned towards a more consultant led service thus necessitating an expansion in consultant numbers. The pressures to accommodate service provision within the restricted hours led to concerns over its effects on training. However, rather than calling for an ease to the work hour restrictions, the report advised that major changes were needed to the traditional experiential model of learning. It found that while patient safety correctly remained a high priority, the service provision relied heavily on trainees to deliver this care and particularly out of hours. The requirement to adapt to the EWTD had resulted in the development of working in shift patterns. This had often had adverse effects on continuity of care and threatened training opportunities. Increased shift work often required increased numbers of doctors and as a consequence there was an increase in the number of vacant positions in the rota. The requirement to fill out-of-hours rota vacancies at short notice, at the expense of elective training opportunities was highlighted as a major cause for concern. In a survey of almost 500 trainees by the Association of Surgeons in Training (ASiT) over half had
reported gaps in rotas and the majority of these were covered by internal locums. Poor rota design often resulted in reduced supervision by senior colleagues thus reducing the learning opportunities previously provided by the same shift.

The report found that shift working in itself had a significant effect on the quality of training even if the rota vacancy issues were addressed. In particular, shift working had often led to a reduction in the time a trainee spends with their trainer. The erosion of the traditional ‘firm’ structure may result in trainers and trainees working shifts at different times of the day or night and reducing the time they spend together as a consequence. The evidence gathered did however suggest that if the service and training delivery adaptations to the EWTD were carefully developed with other changes occurring in the NHS, then the effects of this reduced trainer to trainees’ interaction could be mitigated to a certain extent. Another consequence of the erosion of the firm structure through shift working was the loss of continuity of care for individual trainees. This had detrimental effects both with the loss of recognised learning opportunities themselves but also with the loss of morale by missing opportunities to see the positive effects of a trainee’s hard work. Handovers were identified as a particularly weak event both in terms of training and service delivery. Perhaps the most surprising finding of shift work was that it appeared to be unfavoured by trainees and far from the EWTD improving the work-life balance, shift working had led to a perceived reduction in the quality of life of trainees. Although the number of hours worked had been reduced, the intensity of those hours coupled with the antisocial nature and irregularity of shifts, some with fixed leave, had overall had a bigger impact on trainees’ lives. Circumstantial evidence for this negative impact was provided
by reports of increased sick leave being required by junior doctors following the implementation of the EWTD (McIntyre et al. 2010).

While it was acknowledged that logbooks merely provide an indication of the quantity of training with little direct information regarding the quality, it provided some objective evidence of EWTD effects. Limited data were available, however cardiac surgery reported positive effects of maintenance of operative exposure despite a reduction in work hours (Lim et al. 2006). The report was clear in its message that traditional methods of training based on the experiential and apprenticeship model needed updating. The hours available for training were currently not being used to their maximum potential. Trainees were often performing administrative tasks with little benefit to their training. Then quantity of training in terms of hours worked or the number of procedures performed was being used as a proxy marker of training quality. Training junior doctors is a recognised part of the consultant role, however it was found to have a low level of priority in consultant job plans. Clinical Directors would often focus on service-orientated objectives rather than those for training junior doctors. This is likely due to the service driven quality targets that NHS Trusts are held accountable for, where as similar training quality targets are not available. Furthermore, those consultants that do accept greater responsibility for training are often not appropriately recognised or rewarded for their efforts.

The report recommended the implementation of a consultant delivered service and distinguished this from the less explicitly defined consultant based or consultant led service. This would mean 24-hour consultant presence for direct patient care. With the on-going
increase in the consultant workforce and the planned decrease in trainee numbers, the report called for a redevelopment of the model of training and service delivery. One of the main challenges to delivering high quality training was the role of poorly supervised trainees in out of hours service. With a consultant delivered service, trainees would benefit from improved supervision out of hours thus allowing them to maximise the learning during out of hours service commitments. The report was clear that trainees should be trained within a service environment and not become ‘supernumerary’. There was no appetite for training to be separated from service provision since the report recognised that highly regarded training institutions also delivered high quality clinical care to patients. Graded consultant supervision would allow trainees to gradually taken on more responsibility as their competencies allowed.

Implementation of the EWTD could serve as a catalyst to reconfiguration of service and training provision with a positive impact on the work-life balance of trainees. Positive reports were received on the division of services into emergency and elective clinical pathways and having complimentary emergency focused or elective focused rotas. This was often found to be more predictable and led to more senior supervision. Furthermore, reconfiguring the consultant role could permit those with an interest and the ability to deliver training, to play a greater role in education. Other consultants may wish to play a lesser role in education but a greater role in management or greater focus on service delivery. It was also possible that consultants’ interests and focus may change during their careers. There was acknowledgement that while consultants training in the modern era would be well trained, they will likely have less experience than their predecessors. Consultants would therefore benefit from appropriate
mentoring and support in the early years following their appointment. Importantly it was felt that where services were to be reconfigured, the impact on training should be evaluated before any reconfiguration was implemented.

The increasing consultant numbers and reduction of trainee numbers would affect the balance of service provision and training requirements during reconfiguration of services. This may result in some hospitals not being able to deliver training. Therefore training locations should be chosen by assessing the quality of training a department or hospital can deliver. The report recommended making ‘every moment count’ through focused and where possible, individualised training. Simply being present in hospital for a number of hours was not enough to train doctors. Trainees and trainers needed to maximise learning opportunities in all clinical settings. The report supported co-ordinated, integrated use of simulation and technology to accelerate learning. As part of this recommendation, trainers must develop skills in the use of simulation. Improvements to training would require current best practices to be combined with innovation to deliver better methods of training. This would shift the focus from the number of cases in a logbook to actual competencies achieved and weaknesses that need to be targeted. Structured coaching was promoted as a method to teach competencies and then subsequently expertise.

The importance of training was highlighted throughout the report and to compliment this the report recommended the need to reward and incentivise training. Consultant trainers should undergo some form of accreditation process and be offered support and training to fulfil their
role. Rewarding training need not necessarily be in the form of salary but could take the form of gaining protected teaching time in the form of Programmed Activities (PA). Clinical Excellence Awards could also be encouraged to promote high quality training. Incentivising training at Trust level would allow training to be formally recognised in consultant job plans for those who are nominated as trainers. The commissioning of postgraduate medical education could provide the leverage to develop innovative methods of delivering high quality training. Training quality performance parameters should be included in reviews of Trust management processes. In conclusion, quality control for training would require transparent quality ratings to be developed and monitored for all training units.

1.3.7: The NHS Future Forum

The focus on high standards and high quality was carried forwards by the NHS Future Forum’s work on education and training (DoH 2012). The proposed establishment of Health Education England received strong support. At a local level, the establishment of Local Education and Training Boards (LETBs) was welcomed as an opportunity to develop collaboration between research, education and service delivery.

The forum found evidence that education and training was not prioritised in some organisations. This could be demonstrated in a number of different manners, for instance by non-attendance of management at meetings concerned with education arrangements, or through preventing staff from attending training opportunities. For a Hospital Chief Executive, pressure to deliver clinical services is likely to conflict with their interest in supporting surgical
training. Furthermore the report confirmed the belief that training is not always of the highest quality and in fact some is not fit for purpose. A major concern raised by the forum was the processes in place to address these poor quality training units. The threat of removal of trainees was rarely implemented due to the pressures to deliver clinical services. Concerns were raised that careers in clinical education were not well regarded within the profession and this was contributing to reducing the priority of education within organisations.

Contrasting views were received on the duties of NHS organisations to provide training, with some believing that any organisation receiving NHS funding had an obligation to provide training. However the review’s recommendations were clear that training should only be provided where it could be delivered to high standards. This would mean that not all organisations, departments or clinicians should offer training. Clinicians who have an interest in training should be appropriately supported both in terms of their development as a trainer and in providing enough time to train. Training could be incentivised by rewarding excellence in training through the development of quality premiums by developing an education ‘Commissioning for Quality and Innovation’ (CQUIN) payment. In order to implement this, organisations would need to demonstrate evidence of high quality training through evidence based outcomes metrics. The NHS Future Forum had previously recommended the development of a comprehensive quality assurance framework with explicit educational outcomes. At a commissioning level, services should be commissioned from providers that provide evidence of high quality education and training.
Transparency in funding streams was called for with relation to all forms of training. The forum’s view was that although funding levels were adequate, concerns had been raised that education funding was not always used for training. They recommended a transparent approach to funding and that only high quality training units should be funded. The forum proposed this could be achieved through a tariff for trainees so in effect the training budget would follow the trainees. Transparency would be accompanied by robust accountability to ensure that all training budgets were used for the correct purposes they were intended for.

1.3.8: The Shape of Training Report

As a response to the numerous reports calling for reforms to postgraduate medical training, in 2011 a national cross specialty steering group was set up to identify possible reforms to the structure of postgraduate medical education and training in the UK. The final report (Greenaway 2013) considered the recommendations from previous inquiries into medical education. The sponsoring board highlighted Sir John Tooke’s Aspiring to Excellence report in which he called for flexible, broad based training while integrating service and education with workforce planning. The Shape of Training Sponsoring Board comprised:

- Medical Education England (MEE) and Health Education England (HEE)
- Academy of Medical Royal Colleges (AoMRC)
- General Medical Council (GMC)
- Medial Schools Council (MSC)
- Conference of Postgraduate Medical Deans of the UK (COPMeD)
• NHS Education Scotland (NES)
• Northern Ireland Medical and Dental Training Agency (NIMDTA)
• Wales Deanery

The sponsoring board selected the Vice Chancellor of the University of Nottingham, Professor David Greenaway to lead the independent review. Five key themes were identified for the review to consider:

• Patient needs
• Workforce needs: the balance between specialists and generalists
• Broad approach to postgraduate medical education
• Tensions between service and training
• Increasing the flexibility in training

In addition the review was also asked specifically to examine the transition from medical school to the Foundation Programme. Many stakeholders were engaged including patients, student and trainee doctors, consultants and employers. Following a literature review, nine site visits and five large seminars were conducted as part of the engagement process. Nearly 400 written submissions were received. After considering all the information, the principles and recommendations were discussed in workshops and 54 oral evidence sessions. In total, feedback was received from over 1500 individuals and organisations.

The main message of the recommendations was to provide doctors with a flexible and broad based specialty training rather than focusing on narrower sub-specialist training. Patients wanted to be clear about the seniority and the competency of the doctors treating them.
Formal credentialing of competencies would allow doctors to enhance their practice or transfer competencies to other specialties or subspecialties.

The review recommended supporting transitions during the medical career pathway. Particular attention was paid to the transition from medical student to the Foundation Year 1 post. The feeling amongst the profession that patient safety is compromised during the arrival of the newly qualified doctors in August (Vaughan et al. 2011) was supported by evidence that patients admitted on the first Wednesday in August had a higher early death rate those admitted a week earlier (Jen et al. 2009). The review also heard that there was concern amongst trainees and trainers about how well prepared graduating medical students are to enter clinical practice. In a survey of over 6000 Foundation Year 1 doctors conducted one year after graduation, only 51.4% felt that medical school had prepared them adequately for clinical practice (Goldacre et al. 2014). The review recommended that medical schools must ensure medical graduates can work safely in a clinical role. In agreement with the NHS Future Forum’s recommendations, the review supported limiting postgraduate medical training to high quality training units and called on the GMC to approve and quality-assure these units.

The report continued to support the move towards a more competency based progression. They heard how time was still being used as a proxy measure to confirm competency in many instances. Consultant trainers voiced their concerns over the diminishing relationship they now have with their trainees. The report recommended longer placements to try to improve the supervision of junior doctors through better integration of trainees within the multidisciplinary
teams and by allowing them to build stronger relationships with their trainers. The report recommended an apprenticeship-based arrangement between trainee doctors and their consultant trainers. In particular this would facilitate a more personalised supervision for the trainee. The *Temple report* had stated the traditional apprenticeship model required updating as it largely relied on experiential learning. The *Shape of Training* report highlighted some key elements of a successful apprenticeship-based arrangement. These included training supervisors and supporting them for the role, trainees and supervisors working together on days, nights and weekends, and the adoption of mentoring and coaching techniques. This would require significant resources and in view of this the report confirmed that not all doctor should be trainers and not all units would be training centres. Training should be reserved for units that can demonstrate evidence of high quality training and supervision. The report called for an immediate review into the best methods to quality assure training environments and subsequently restrict training to high quality units.

The *Shape of Training* report was published in October 2013 and whether the recommendations will be implemented remains to be seen. After years of subspecialisation which is particularly well established in general surgical specialist training, it would be a paradigm shift to revert back to broad based general surgical training. Furthermore, while the report was categorical in its desire not to see a ‘sub-consultant’ post established, fears will remain that this will be an inevitable consequence of subspecialty training being moved to the years beyond CCT.
1.4: Focusing on Quality in Education

The Shape of Training Review’s explicit focus and continued support of restricting training to centres and trainers with a demonstrable track record of high quality training was consistent with previous reports. The concept was first introduced in Aspiring to Excellence, with the recommendation to nationally centralise higher specialist training commissioning and award Trusts on a competitive basis. It recommended the development of performance measures to facilitate commissioning, incentivising and reporting of training. In the NHS Next Stage Review, efforts to improve the quality of training included mandatory training provision for educational supervisors and regular performance review. Accountability of training quality would be enhanced through educational commissioning by means of a tariff-based system assigned to individual trainees. The Temple report suggested that educational commissioning could provide leverage for development and innovation in methods of delivering high quality training. It described the need for transparency of any quality ratings developed and the need for longitudinal monitoring of training quality. The NHS Future Forum proposed a CQUIN payment for education as a means of incentivising excellence in training. It further supported the calls for transparency of funding streams through the tariff for trainees. The need and desire to assess and improve training quality has been established, the challenge remains how to implement these principles.
1.4.1: Assessment of Quality

Quality assurance processes for training were mentioned in *Unfinished Business* but few details of how this should be achieved were provided. In *Aspiring to Excellence*, the market forces rationale for why quality assessment was necessary in surgical education was explained. Protected health research funding was facilitated by the established methodologies through which research output could be assessed. Research output was measurable thus providing a quantifiable measure of research quality. Similarly financial, clinical and service delivery targets could be incentivised since there were measurable outcomes to justify resource allocation.

There was no established methodology to assess quality of postgraduate medical education.

There appears agreement that not all consultants and hospitals may be suitable for surgical trainee placements. However, in order to enable hospitals and consultants to demonstrate their suitability and commitment to high quality training, it is first necessary to develop a method for measuring training quality that will enable hospitals to provide the required evidence. The challenge arises in defining high quality surgical training. Without establishing this definition, training quality cannot be measured. The *Temple* report made some attempt to seek views on the definition of high quality training. It went on to highlight the lack of established quality measures that evaluated the outcomes of training.

In response to the Future Forum on education, the Royal College of Surgeons of England (RCSEng) highlighted that(*RCSEng 2011*) education and training of surgeons:
'Is fundamental to the delivery of quality and therefore patient safety both now and in the future' and recommend ‘the introduction of independent specialist scrutiny of training quality……clear quality indicators to ensure that training opportunities are maximised.’

The RCSEng describe monitoring tools, which include deanery visits to training providers, online trainee surveys and trainee logbook analysis. The Joint Committee on Specialist Training (JCST) have developed quality indicators (QIs) in an attempt to assess the quality of training placements(JCST 2012a). Similar indicators can also be inferred from the American Board of Surgery (ABS) and the Accreditation Council for Graduate Medical Education Surgical Residency program requirements(ACGME 2012). In the USA the Residency Review Committees are charged with the responsibility of accrediting training institutions against predefined criteria.

The logistics of conducting deanery inspections to training institutions and annual applications for reaccreditation of training status can be unwieldy. Nevertheless, some form of assessment is essential to ensure high quality training is maintained. The electronically administered JCST trainee survey was developed to reflect their quality indicators and measure them. While encouraging trainee feedback is a welcome step in the right direction, this evaluation of quality relies solely on the subjective views of trainees. There is evidence that trainees and trainers demonstrate significant differences in their definitions of service and education(Sanfey et al. 2011). Therefore trainees may provide an unbalanced assessment as they may view as service what trainers feel is training and vice versa. The details of how these quality indicators were developed are not clear and the indicators themselves are somewhat vague e.g. trainees
'should have the opportunity to undertake a wide range of operations in elective and emergency General Surgery'. The JCST survey answers are available to Training Program Directors, Heads of School, Specialist Advisory Committee (SAC) Chairs and SAC Liaison members.

The assessment of quality may seem novel in terms of surgical education, however in other domains it is well established. If we look outside of medicine to other educational institutes such as universities, the assessment of quality of education is multifactorial including ratings for faculty, academic reputation, etc. Their methodologies are published online and are transparent (USN 2012). Within surgery, the evaluation of quality in clinical outcomes has been a strong driver for raising clinical standards. The introduction of the National Surgical Quality Improvement Program was hailed as the most comprehensive comparative clinical outcomes assessment program with the objective of enhancing the quality of surgical care (Khuri et al. 1998). There is evidence that the use of NSQIP can improve the quality of clinical care (Compoginis and Katz 2013; Hall et al. 2009). In the UK in order to be officially recognised as an Academic Health Science Centre (AHSC) a unit needs to demonstrate excellence in research, patient care and education. The introduction of Dr Foster league tables (DFRL 2011) and the evolution of the Research Assessment Exercise (REF) have undoubtedly encouraged institutions to invest more resources into improving clinical and research outcomes to raise the performance in these nationally reported assessments. Introducing an objective assessment tool to evaluate the quality of surgical training programs could act as an incentive to renew the focus on surgical training outcomes both in the UK and globally.
1.4.2: Innovation to Improve Quality

There is renewed interest in postgraduate medical education in England (Ovseiko and Buchan 2011). This moment presents an exciting opportunity to improve the quality of surgical training through innovative and evidence-based recommendations. The volume outcome relationship (Birkmeyer et al. 2002; Birkmeyer et al. 2003) has certainly had a profound effect on the distribution of surgical services and thus indirectly on training. However, it is not clear whether these conclusions can be extrapolated to surgeons in training. It would be wrong to assume that volume is the sole indicator of high quality surgical training. While exposure through time and consequently through caseload, undoubtedly has a significant effect on surgical training, there is a need for pragmatism and to explore innovative methods to optimise the quality of training without simply relying upon increased caseload. In the Chief Medical Officer’s annual report in 2009 (L. Donaldson 2009), one of the five themes was assigned towards the benefits of simulation training. The recommendations were to fully integrate simulation training within all stages of clinical training programmes. In particular, the report called for a skilled faculty to maintain high standards of training and for simulation based training to be funded and valued by NHS organisations.

The traditional Halstedian apprenticeship model was based largely on experiential learning through practice on patients. This is no longer acceptable to the public and with the global pressure to reduce training hours; it is no longer feasible at an organisational level. In fact there is evidence that despite a reduction in absolute numbers of patients seen, redesigning an internal-medicine residency program can have demonstrable training benefits (McMahon et al. 2008).
This prospective controlled cohort study covered a 12-month period and almost 4000 patients, following the implementation of an experimental inpatient medical service with reduced resident workload. Despite seeing fewer patients, trainee satisfaction and time for educational activities increased in the experimental group with no difference on adherence to standards for quality of inpatient care. As a craft specialty, surgery is often perceived as being more reliant on experiential learning to ensure sufficiently large numbers of operative cases. However, there are reports that operative exposure can be maintained despite restrictions to working hours (Lim et al. 2006). This study reported greater numbers of procedures performed by trainees following the implementation of working hour restrictions. The success was attributed to the innovative implementation of a team approach to surgical training where surgical trainees were allocated to operating theatres with a case mix appropriate to their level of experience. The authors commented that the success of their innovative training model stemmed from a culture that was firmly committed to surgical training. This resulted in a greater proportion of training cases made available to the junior trainees.

The aforementioned government reports into postgraduate medical education coupled the desire to reliably and robustly assess the quality of training, with the search for innovation to improve training quality. The NHS Next Stage Review (DoH 2008a) mentioned the commissioning of research to identify high quality educational practice. The Temple report’s message was that the hours worked were not being used to their maximum potential for training. It called for current best practices to be combined with innovation to deliver better methods of training. The report suggested that the commissioning of medical education could
provide the leverage to develop innovative methods of delivering high quality training. Structured coaching was promoted as a method to teach both competency and subsequently facilitate the acquisition of expertise. The Shape of Training review also highlighted the adoption of coaching and mentoring techniques as a key element of a successful apprenticeship-based arrangement, rather than the traditional experiential apprenticeship model. The concept of coaching and its potential use with surgeons has stimulated much discussion following an article in The New Yorker written by Professor Atul Gawande (Gawande 2011). In it Professor Gawande strikes the analogy between a top-level professional tennis player receiving coaching and the fact that this technique is seldom used in surgery. The use of structured coaching could indeed provide an innovative method to train surgeons. The consultant trainer would be the natural coach for a surgical trainee. However innovation would be required to ensure sufficient time was available for coaching to be viable. The Shape of Training review heard from consultants who were concerned at the seemingly diminishing relationship they appeared to now have with their trainees. The Temple report highlighted the missed training opportunities during out of hours service provision commitments of trainees. The introduction of shift working patterns to accommodate the EWTD had often led to a reduction in time a trainee spends with their trainer. Video based coaching would be one innovative method of enhancing training quality. It may also be an innovative method of providing the graded consultant supervision supported by the Temple report. This could allow trainees to take on more responsibility as their competencies allowed, while still allowing them to learn and progress from competence to expertise. While video-based coaching of surgeons
has been qualitatively investigated (Hu et al. 2012), further empirical investigation is required to see if it would be effective.

Transitions to practice are another potential area to focus on innovating to improve training quality. These can be at the interface of graduation from medical school, the progression from core to higher specialist surgical training and the ultimate progression from trainee to independent practice as a trained consultant surgeon. This latter transition has been the focus of much attention from the American College of Surgeons (ACS) recently. The growing proportion of trainees that seek fellowship training evidences the challenges of the transition to independent practice. Over 80% of graduating USA general surgery residents are now pursuing a fellowship (Lewis and Klingensmith 2012). In a survey of Fellowship Program Directors in North America, 42.7% of respondents felt that on arrival to the fellowship, graduating residents were unable to independently perform 30 minutes of a major procedure (Mattar et al. 2013). However upon completion of the fellowship, 89% of respondents felt that the trainees were sufficiently prepared for independent practice. The ACS has developed a Transition to Practice Fellowship in General Surgery Program to focus specifically on graduating surgical trainees (ACS 2013).

In the UK, the Shape of Training review was asked to look in detail at the early career transition from medical school to the Foundation Programme. They reviewed the objective evidence that patient safety was compromised during the transition of junior doctors at the start of August (Jen et al. 2009). This nationwide study of almost 300,000 patients used 9 years of the
Hospital Episode Statistics (HES) administrative dataset in England, to examine the mortality of emergency admissions on the first Wednesday of August and compared it with the previous Wednesday. The first Wednesday in August is when graduating medical students start their careers as qualified junior doctors. The study demonstrated a significant 6% higher odds of death in the week following the first Wednesday in August when compared with the previous week. This evidence supported the concerns of the profession about the preparedness of newly graduating medical students (Vaughan et al. 2011). Only just over half of Foundation Year doctors rated themselves as adequately prepared for clinical practice (Goldacre et al. 2014). The NHS Future Forum had similarly heard concerns about the preparedness of medical graduates to commence their clinical duties. They had suggested reintroducing the opportunities for incoming Foundation Year 1 doctors to shadow their incumbent position holders. This was seen as a simple strategy to start addressing the issue. For a more comprehensive approach, the Tooke report identified an opportunity to explore whether ‘greater and more challenging service experience can be gained during the later stages of the undergraduate programme.’ The Temple report had highlighted that where appropriate a simulated environment should be used to learn clinical skills rather than on patients. This combination of the profession itself feeling ill prepared at graduation together with objective evidence worsened patient safety highlights this transition form medical school to clinical practice as an opportunity to innovate high quality teaching. The use of intensive simulation skills boot camps would seem an obvious innovation to address these challenges. This has shown some encouraging results in early studies although these have largely been using self reported confidence as the outcome measures.
The search for innovation and evidence should not be restricted to the UK as this is a global issue, enlisting considerable effort and investment to enhance surgical practice. There are marked global variations in surgical training programs. The most marked of these are differences in the lengths of training programs (see Figure 1c). In some countries surgeons are trained in half the time it takes a surgeon to be trained in another country. Figure 1c also demonstrates variation in the timing and number of competitive selection points for surgical trainees. Furthermore, while the concept of ‘run through’ training was novel when introduced to the UK with the MMC reforms, it is the standard practice in some countries such as the USA. Despite these variations, the common end product of these programs is a trained surgeon capable of independent practice. This is supported by evidence to suggest that the knowledge and technical skills of surgeons trained in different countries under their respective training programs are equivalent (Schijven et al. 2010). It follows that if variations exist and yet the end product is similar despite shorter training, then innovation within curricula have the potential to enhance the quality and efficiency of surgical training. This thesis aims to identify these variations, develop a method to measure them and subsequently develop and pilot training interventions that can improve the quality of surgical training.
Figure 1c: Postgraduate Competitive Entry Points Into Surgical Training and Mandatory Examinations

Examinations
1.5: Hypotheses

- There are variations in global surgical training programs.

- Quantity of training measured by time or caseload is not directly proportional to the quality of training.

- Quality of training can be measured.

- Training improvements can be made without increasing the quantity of training but by improving the quality of training.
Chapter 2: Global review of surgical curricula

The following chapter was published as:

2.1: Introduction

In order to identify indicators of high quality surgical training, it is first important to identify the variations in surgical training programs. This chapter will examine the differences between the current surgical training programs globally.

In the past decade there has been a global drive to standardise national surgical training curricula and align these with the paradigm shift towards competency-based training. The objective of this chapter is to compare surgical curricula from around the world and to identify variation. Identification of variations will permit evaluation of the strengths and weaknesses of surgical curricula. This could provide information on methods to improve curricula to efficiently and effectively deliver high quality surgical training.

2.2: Methods

A convenience sampling method was employed to conduct an online electronic search for nationally recognised English language general surgical curricula. The search resulted in curricula of Australia, Canada, Hong Kong, the UK and the USA from the websites of their respective national surgical colleges(CSHK ; GSA ; ISCP ; RCPSC ; SCORE). An initial qualitative evaluation of the development and structure of each curriculum was performed.

To enable comparisons of the curricula using standardised criteria, a search was performed for an established set of objectives to evaluate the curricula against. The search identified the UK
General Medical Council’s (GMC) ‘Standards for curricula and assessment systems’ (GMC 2010). This consists of 17 standards grouped under 5 headings: Planning, Content, Delivery, Outcomes and Review. Each GMC standard consists of a number of subdivisions. For the purpose of this chapter, each curriculum was assessed against each standard and given a rating based on how well the curriculum met the GMC standard;

- Meets the standard
- Meets the majority of description of the standard
- Meets some of the description of the standard
- Either does not state the information required or does not meet the standard.

In order to meet the majority of the standard the a priori definition was that the curricula had to meet 50% or more of the subdivisions within the description of the standard.

The results are presented as two sections with an initial overview of each curriculum in the first section. This is then followed by comparisons across the curricula with respect to each of the GMCs standards.
### 2.3: Results

The standard surgical training pathway in the countries examined is displayed graphically in Figure 2a.

**Figure 2a: Standard Surgical Training Pathway**
2.3.1: Australian Surgical Curriculum

General Surgeons Australia (GSA) collaborates with the Royal Australasian College of Surgeons (RACS) to deliver training in Australia through the SET Training Program. The GSA curriculum is available online through the GSA website in a number of separate documents covering curriculum outlines, full curriculum modules, SET requirements, SET competencies and a training regulations handbook. The curriculum outline is in a modular structure outlining the rationale and objectives, suggested reading, a description of the learning opportunities and methods and finally information on how the module will be assessed. For each module, this is then followed by a table of two columns that describes the knowledge of anatomy/physiology/pathology expected for each disease process at clearly identified points in the training pathway.

The full GSA curriculum(GSA) consists of 14 technical and 6 non-technical modules with a similar structure to the curriculum outline modules. The names of both the developers and reviewers of each module are clearly stated at the start of each module. Non-technical modules include a self-assessment section and are framed around the CanMEDS roles (as described below). The technical modules again employ a table format to state competencies expected at clearly identified SET training levels. The competencies are described under the following headings:

- Medical Expertise
  - Anatomy, Physiology, Pathology
- Judgement/Clinical Decision Making
○ Clinical assessment
○ Investigations
○ Principles of management

• Technical Expertise
  ○ Operative management – does
  ○ Operative management - knows

The GSA SET 1 Competencies (Anthony et al. 2008) bear significant similarities to the CanMEDS roles. The methods of assessment are described and a list of index procedures is provided. The GSA SET 1 Requirements (Anthony et al. 2009) outline the mandatory clinical experience, skills courses, in-training assessments and examinations. The requirements to ensure progression from SET 1 to SET 2 are explicitly stated.

The GSA training regulations handbook (GSA 2011) provides a comprehensive overview of the structured 5 year training program. The structured assessment program bears some similarities to the UK’s ISCP. Workplace based assessments include Direct Observation of Procedural Skills in Surgery (DOPS) and Mini-Clinical Examination (Mini-CEX). Maintaining an accurate logbook of surgical operations is mandatory and a progressive increase in the proportion of primary operator experience is expected. Minimum numbers of operative cases necessary are specified. Trainees are required to register with a conjoint committee that oversees endoscopy training. Research activity is a mandatory part of the program. There are a number of courses that are
mandatory and these are made explicit. Eligibility criteria for the fellowship examination and completion of education and training are provided.

2.3.2: Canadian Surgical Curriculum - CanMEDS

The CanMEDS Physician Competency Framework (Frank 2005) was developed by the Royal College of Physicians and Surgeons of Canada along with multiple stakeholders including trainees, educators and patients. Development commenced in 1993, and following a pilot phase, the framework was implemented from 1997-2002. A fourth phase of faculty development began in 2002 and an updated version was published in 2005. The central theme is a competency-based framework that describes the principle generic abilities of physicians oriented to optimal health and health care outcomes. These are presented as the CanMEDS cloverleaf incorporating the 7 CanMEDS roles:

- Medical expert (central)
- Communicator
- Collaborator
- Health Advocate
- Manager
- Scholar
- Professional

The structured framework defines and describes each of the seven CanMEDS roles and lists the key elements and key competencies. The generic framework is complemented by documents specific to the General Surgery Curriculum (RCPSC 2010d, 2010a, 2010b, 2010c). Objectives are
given both for generic surgical foundations training and training in General Surgery. Specialty specific competencies cover knowledge, assessment, interventions and technical skills. There is a strong emphasis on competency as oppose to minimum numbers of procedures. Minimal details are provided regarding procedures.

A minimum of 2 years of surgical foundation training is mandatory which must include a block of critical care and initial trauma management. These years can be incorporated into the 5 years of specialty training in General Surgery. The rotational and examination requirements for this higher surgical training are explained along with the format of the Final In-Training Evaluation Report (FITER)(RCPSC 2010e). A research component is essential to complete the program.

Standards for the accreditation of residency programs are specified in both generic(RCPSC 2011) and general surgery specific documents(RCPSC 2010f). Programs must be under the direction of a Canadian University Medical School. Clinical services that are used for teaching must be organized to promote their educational function. The General Surgery specific standards give general guidelines for the objectives, structure, resources and content within residency programs. Programs lacking specific resources can be accredited so long as formal arrangements have been made to send residents to other accredited residency programs to fulfil the deficient area. Guidance is provided regarding the methods employed to evaluate residents within the program but no formal framework for in training assessments is in place.
2.3.3: Hong Kong Surgical Curriculum

The Hong Kong surgical curriculum is developed by The College of Surgeons of Hong Kong (CSHK). A 2-year basic surgical training program is followed by 4 years of higher surgical training in General Surgery. The revised curriculum for Basic Surgical Training (BST) (CSHK 2013) was published in 2013. Upon completion of BST the trainee is expected to be competent in 5 of the 7 CanMEDS roles: Medical expert, Communicator, Collaborator, Scholar, Professional. During BST, rotations include various surgical specialties and declaration of interest is on a voluntary basis only. Those trainees who do not declare an interest are placed in General Surgery. The details about lengths of rotations and timing and eligibility of examinations are outlined. The modular curriculum is accompanied by workshops and courses that are subdivided into those that are mandatory and those that are relevant. Each surgical subspecialty has a one-page summary of required competencies that follow the CanMEDS roles closely and includes lists of index procedures. Index procedures are divided into surgical or endoscopic skills that require assessment and those that the trainee should receive exposure to. Generic competencies follow the CanMEDS structure. The structured assessment system uses Mini CEX (Clinical Evaluation Exercise), DOPS (Direct Observation of Procedural Skills) and DOPS in Endoscopy to assess competencies in the workplace in addition to a 3 part written and clinical examination. Supervisor assessment forms are standardised with numerical scoring of the trainees on clinical and technical skills, academic performance and attitudes. The trainees are required to keep all of their assessments in a Logbook along with detailed operative records. These are then inspected prior to entry to higher surgical training.
Higher surgical training is structured with clear entry criteria and regular assessments. The trainee’s mentor conducts a face-to-face assessment in addition to a logbook review every 3 months. Twice a year the trainee submits their logbook and two mentor assessment forms for review by the specialty board. There is a conjoint exit examination with the Royal College of Surgeons of Edinburgh. All training posts require approval from the College of Surgeons of Hong Kong with a recommended 1:1 trainer to trainee ratio. Lists of hospitals with approved programmes for higher surgical training are provided. Trainees are required to rotate out of their parent training centre for a minimum of one year.

There is a strong emphasis on volume of cases with index operations listed and specified minimum numbers of procedures provided. A ‘major’ operation is defined as ‘cholecystectomy and above’ and a minimum of 100 per six months for each trainee is recommended. In addition to the absolute numbers, a graduated increase in the proportion of cases the trainee performs as the primary surgeon is specified in a table of ‘Flag criteria’ for logbook summary reports. Minimum numbers of index procedures to be achieved by the end of HST are specified. The required competencies of a graduating trainee are described under 9 subheadings which bear resemblance to the CanMEDS roles(CSHK 2004).

### 2.3.4: UK Surgical Curriculum - ISCP

The UK surgical curriculum has undergone redevelopment with involvement from the Royal Colleges of England, Edinburgh, Glasgow and Ireland resulting in the Intercollegiate Surgical Curriculum Programme (ISCP). There were four phases of development that included pre-pilot
and pilot phases. The Intercollegiate Surgical Curriculum Programme (ISCP) was implemented in August 2007 whilst the fourth review and evaluation phase continued until March 2008. It is an online curriculum (ISCP) with a detailed overview, syllabus and a structured assessment and feedback system. The website provides an online database for assessments and in addition features a link to the trainees’ online operative logbook.

Following 2 compulsory years of foundation training, surgical training is divided between 2 years of core surgical training (CT1-2) in a broad range of surgical specialties followed by competitive entry into 6 years of specialist General Surgery training (ST3-8). The specialist training years are further subdivided into intermediate (ST3-6) and Final (ST7-8) years although there is no reselection procedure at this point. The standard rotations are 6 months with regular emergency on call commitments. Training is limited to a maximum of 48 hours per week in line with the European Working Time Directive (EWTD). The curriculum aims to train general surgeons capable of independent practise at consultant level. However trainees are expected to develop a subspecialty interest such as Colorectal or Transplantation etc.

The syllabus (ISCP 2010) outlines the required competencies of a trained surgeon to receive their Certificate of Completion of Training (CCT). This is divided into elective, emergency and special interest topics with lists of index procedures provided. All stages include generic professional behaviour and leadership skills. Modules are described under the following subheadings: Objective, Knowledge, Clinical skills and Technical skills, with assigned competency standards stratified by stage of training.
There is a clear emphasis on competency-based training with details of the entry requirements for specialist general surgery training and a structured series of workplace based assessments (WPBAs) that are recorded online. Training progression is demonstrated by acquisition of the levels of knowledge and clinical and technical skills determined for each stage. At an annual review of competence progression (ARCP), a combination of WPBAs and assessor feedback determine trainee progression to the next stage. Gastrointestinal (GI) surgeons are additionally expected to achieve competency in endoscopy as set out by the Joint Advisory Group on GI endoscopy.

2.3.5: USA Surgical Curriculum - SCORE

The training program comprises a 5-year General Surgery residency. Work hour restrictions enforce a maximum 80 hours per week. A national General Surgery curriculum was originally developed in 2004 by the General Surgery Residency Committee of the American Board of Surgery in response to calls for a better defined curriculum in General Surgery residency training. The Surgical Council on Resident Education (SCORE) was formed in 2006 as a non-profit consortium made up of the following 7 US surgical organisations:

- American Board of Surgery (ABS)
- American College of Surgeons (ACS)
- Association of Program Directors in Surgery (APDS)
- Association for Surgical Education (ASE)
- Residency Review Committee for Surgery of the ACGMA
• Society of American Gastrointestinal and Endoscopic Surgeons (SAGES)

Their objective was to develop a standardised national curriculum for General Surgery residency training. The first SCORE curriculum outline was developed using an expert consensus process and published in 2008 and following annual reviews; an updated version is now available for 2014-2015. While it was developed as a national curriculum, its use is not compulsory and involves a subscription fee. The current subscription fees are set at $500 for the annual program license and an additional $125 per-resident fee.

The SCORE outline(SCORE 2011) is the first step in curriculum development and is not a full curriculum which would include educational content, teaching materials and assessment. The outline has been approved by the ABS and SCORE as the basis for full curriculum development. It is intended to be used in conjunction with the SCORE General Surgery Resident Curriculum Portal(SCORE), an online resource that provides teaching materials in support of the curriculum. The curriculum outline consists of 28 organ system-based categories presented as lists with subdivisions indicating the level of competence required as shown below:

• Disease/conditions
  o Broad: provide comprehensive management
  o Focused: diagnosis and stabilisation, not expected to provide comprehensive management

• Operations/Procedures:
  o Essential – Common: competency required, attainable primarily by case volume.
Essential – Uncommon: often urgent, competency cannot be obtained by case volume alone

Complex - generic experience required but competency in individual procedures not expected

The SCORE Portal is accessed via password authentication. Topics will be assigned a module on the portal with objectives, reading materials and questions. The SCORE Portal is currently operational but still under development, in particular the assessment system is not fully developed. The implementation of the collaborative ACGME and ABS General Surgery Milestones Project in 2014 may well help to address some of these issues (Nasca et al. 2012).

The SCORE Portal is expanding into a large resource for surgical trainees with selected reading materials from textbooks, access to selected radiology and anatomy images, operative videos and evidence based reviews in surgery. The website is interactive allowing creation of assignments and modules which can be assigned to residents and is used to send out a weekly ACS surgery curriculum as well as providing access to practise questions for the ABS In Training Examination (ABSITE).
### Table 2a: Comparison of Surgical Curricula Against GMC Standards

<table>
<thead>
<tr>
<th>GMC Standards</th>
<th>UK - ISCP</th>
<th>USA - SCORE</th>
<th>Hong Kong</th>
<th>Canada</th>
<th>Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose and development</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment system</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General/Professional/Specialty specific</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment system</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Model</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Experiences</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment system</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessor</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competency standards</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review and updating</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resources</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lay and patient involvement</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equality and diversity</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Key**
- Green: Meets description of standards
- Blue: Meets majority of description of standards (>50%)
- Yellow: Meets some of the description of standards (<50%)
- Red: Either not stated or does not meet description of standards

98
2.3.6: Comparison with GMC Standards

Table 2a provides a summary of the comparison of the surgical curricula against GMC standards and these are expanded on below.

2.3.6.1: Planning

This section consists of two standards referring to the curriculum development process and the development of the assessment system. The ISCP and Australian curricula meet both with clear descriptions of trainee involvement in the development and a well described structured assessment system. The Australian curriculum names both the developers and reviewers for each module. The SCORE curriculum provides minimal details on the delivery of the curriculum and is yet to develop an assessment system. Neither the Hong Kong nor the SCORE curricula provide clear evidence of trainee involvement in development. The Hong Kong curriculum was developed by the Council with advice from the education subcommittee to define surgical training requirements; however no further development details about how consensus was reached are provided. In contrast the CanMEDS curriculum provides an in depth outline of its development. Whilst guidelines regarding assessment, feedback format and frequency are provided, there is no standardised assessment framework for continuous assessment.

2.3.6.2: Content

This section contains two standards the first of which refers to the content of the curriculum in terms of the scope of the syllabus, the training structure, methods of learning and a second
standard referring to the structure of the assessment systems both in the workplace and in terms of examinations. The ISCP, Hong Kong and Australian curricula meet both of these GMC standards with structured training programs covering both generic and specialty specific areas, with clearly defined entry and progression requirements. In addition, their respective assessment systems contain clear descriptions of the numbers and types of workplace based assessments and the national examinations required. The SCORE curriculum does not provide a detailed description of the training structure. It focuses on patient care competencies and does not provide generic professional competencies. The assessment system has not been developed. CanMEDS provides minimal details regarding methods of learning, teaching, feedback and supervision. The content area is not benchmarked to identifiable stages of training other than to foundations and at completion. In addition no framework of workplace-based assessments is offered, save general guidelines. The Hong Kong curriculum clearly benchmarks operative volume to training level.

2.3.6.3: Delivery

The four standards of this section cover curriculum implementation, the learning model, learning experiences and the delivery of the assessment system. All the examined curricula meet the standards for the learning model and learning experiences with a balance of workplace based experiential learning combined with ‘off the job’ education in the form of recommended taught courses. The SCORE curriculum does not describe the means of ensuring curriculum coverage. The remaining curricula provide detailed descriptions of the management structure to implement the curriculum including descriptions of the roles of programme
directors, trainers and trainees. Both the SCORE and CanMEDS curricula lack description of methods for on-going assessment. All curricula provide minimal details of the development, reliability and in particular, the evidence base for assessment methods.

2.3.6.4: Outcomes

This section describes five standards covering supervision, assessors, feedback, competency and documentation. The SCORE curriculum does not detail the mechanisms of supervision, criteria for recruitment of assessors, feedback assessments, and how these should be documented. Competency standards are not explicitly mentioned in the SCORE curriculum although the examinations are set by the American Board of Surgery who have detailed descriptions of the standards available on their website. The remaining curricula set out supervision standards with a largely 1:1 ratio of supervision for trainees. In Australia, supervisors are nominated by the specialty training board and approved for a three-year term. Training courses for supervisors are also available. Across all curricula equality and diversity training for assessors is an area that could be developed together with greater details regarding the recruitment of assessors. CanMEDS provides little detail regarding its assessment and feedback system. The ISCP, Hong Kong and Australian curricula meet the feedback standards that include outcomes of assessments and regular monitoring of trainee progression within a well-documented framework. In addition, in Hong Kong the trainee’s mentor conducts a face-to-face assessment as well as a logbook review every 3 months.
Competency standards are met to a certain extent by ISCP, Hong Kong, Canadian and Australian curricula with national examinations and clearly stated standards for completion of training. However, ISCP does not clearly describe the procedure for the right of appeal for trainees while the others do. The Hong Kong and Australian curricula do not detail how pass/fail decisions are reached and how borderline candidates are dealt with. Competency standards are not explicitly mentioned in the SCORE curriculum. All the examined curricula do not describe reasons to choose pass/fail or rank ordering.

The ISCP documentation is standardised and web based, facilitating both accessibility and transferability and therefore meeting the standards for documentation. Hong Kong and Australia have standardised documentation however the paper-based nature limits the transferability and accessibility. The SCORE and CanMEDS curricula do not have standardised documentation of workplace based assessment systems. However, with the SCORE portal, a web based documentation system is under development.

**2.3.6.5: Review**

The final section covers four standards covering updating schedules, resources, lay and patient involvement and equality and diversity. Hong Kong does not give details of curriculum review and updating while the SCORE curriculum acknowledges that it remains under development and is therefore continually being updated. The remaining curricula are monitored on an annual basis. Patient and lay involvement in curriculum updating appears to be limited or absent with the exception of CanMEDS. All examined curricula appear well resourced with facilities both for
trainees and trainers and the appropriate infrastructure to support the delivery of the curriculum and assessments. The online nature of ISCP makes it very accessible and transferable. The resources for assessment of trainees in the SCORE curriculum remain under development but are not currently operational. While the assessment system is not detailed in CanMEDS, the resources and infrastructure are in place.

Lay and patient involvement in development and implementation of assessments is not explicitly mentioned in the majority of curricula. The ISCP describes the involvement of trainees, educationalists and surgeons but no patients. Development of the CanMEDS framework involved input from surveys of patients and research on patient needs combined with public focus groups. Although some curricula mention administrative staff and their views being taken into account, there are no explicit details of laypersons being permitted to act as assessors or examiners. All curricula accommodate a range of learning and teaching methods to respect the needs of patients and colleagues. However they do not specify details of equality and anti-discriminatory policies that local education providers, deaneries and colleges must have in place.

2.4: Discussion

This chapter’s evaluation of general surgical curricula highlights the global variation between surgical training programs, most notably in the length of training. Postgraduate general surgical training is completed within a minimum of 5 years in the USA and Canada compared with a
minimum of 10 years in the UK. Work hours are restricted to 48 hours per week in the UK whilst they are restricted to 80 in the USA and remain unrestricted in Hong Kong, Australia and Canada. Even allowing for the extra hours per week compared to the UK, the total number of hours of training are less in the USA at 19,200 compared to the UK’s 23,040 hours in total (assuming 48 weeks of training per year at the maximum permissible weekly hours limit). Despite this variability, when their respective surgical curricula are compared using objective criteria, with the exception of the USA’s SCORE curriculum, there is little to differentiate the others.

The strength of the Australian curriculum is its transparency; not only are the requirements for progression made explicit but in addition, the processes for grievance and appeals are well described. A very clear set of objectives and timeline of requirements at each stage of training is provided. Workplace based assessments assess competency, however volume of procedures still form a major part of the assessment at each stage. One potential area for improvement would be to create an online database to hold the documentation therefore making it more accessible.

The CanMEDS curriculum provides a very comprehensive set of guidelines for surgical residency training with emphasis on competence rather than volume. However minimal explicit details are provided particularly regarding index procedures and in-training assessments. The main strength is its consistent, logical structure and the overall concept of developing a surgeon who
has multiple roles in addition to those of a clinician. The CanMEDS curriculum has evolved into a benchmarking reference globally for surgical curriculum development.

The Hong Kong curriculum appears organised and the assessment system in particular bears similarities to the UK system. The emphasis on minimum volumes of operative cases seems contrary to the principles of competency-based training. However volume of cases is not used to directly assess the trainees, as they must still prove their competency through assessments. In effect the minimum numbers coupled with the requirement for progressively increasing proportions of operations where the trainee is the primary surgeon may serve as a quality indicator for the surgical training centre. Whilst this concept can be inferred from other curricula, it is in its most explicit form in the Hong Kong curriculum.

The ISCP consists of a well-structured and detailed syllabus that acknowledges competencies will be attained at varying levels of training and creates a framework upon which trainees and trainers can model training progression. The incorporation of a detailed and structured assessment system and online documentation makes this the most comprehensive curriculum of those examined.

The SCORE curriculum is a well-structured outline of a curriculum. However, while it is intended to be used as a national curriculum, its use is not mandatory and requires an annual subscription fee of $500 per program and $125 per-resident. It requires incorporation of structured assessment resources to provide a comprehensive curriculum. It is under continuing
development and would benefit from integration of information currently available through its various member organisations. In particular, competency standards and assessments are currently set and managed by the ABS, which also provides detailed information regarding training program requirements. Furthermore 2014 has seen the introduction of the General Surgery Milestones Project that is a collaborative project between the ACGME and the ABS as part of the Next Accreditation System. The Milestones Project will introduce semi-annual data collection with respect to resident performance evaluation (ACGME. and ABS. 2014). For each competency being evaluated, residents will be assigned a milestone level ranging from ‘Critical Deficiencies’ through levels 1 to 4. Level 1 will indicate the level of an incoming resident, while Level 4 will indicate the graduation target. Integration of SCORE and the Milestones Project would likely improve the SCORE curriculum’s performance against GMC standards as summarised in Table 1. One method of conducting this review would be to include information from the multiple member organisations of SCORE. However this review was intended to be a formative process to identify areas for development rather than to create an international ranking of surgical curricula, therefore this method was not chosen. While the SCORE curriculum will undoubtedly improve through further development in certain areas, the SCORE Portal is already developing into an extensive online teaching resource for learning materials, which is unique amongst the other curricula.

It is evident that no single surgical curriculum is ideal. In fact it would be a challenge even to gain consensus on what the theoretical ideal curriculum would be. There is potential for exams to drive curriculum development rather than the other way around and it is first important to
define the ideal curriculum. The different strengths and weaknesses of the curricula examined demonstrate that there is clear potential to improve them all. However, it is possible to construct a surgical curriculum that meets almost all of the GMC standards by identifying the strengths of the curricula analysed in this chapter. This is the pragmatic design of an ideal surgical curriculum constructed from the strongest aspects of current global practice. Table 2b sets out the GMC standards and alongside each is a brief description of the best method to meet this standard as demonstrated from the curricula reviewed in this chapter. Further studies would be required to evaluate whether these changes could potentially lead to higher quality surgical training and potentially improved clinical outcomes.
Table 2b: The ideal surgical curriculum constructed from the best components of the curricula examined

<table>
<thead>
<tr>
<th>GMC Standards</th>
<th>The ideal model - based on the best example from the curricula examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td></td>
</tr>
<tr>
<td>Purpose and development</td>
<td><strong>CanMEDS</strong>: Developed with input from Fellows, experts, lay experts, literature and public focus groups. A Delphi methodology was used to reach consensus. Surveying two cohorts of fellows subsequently validated the findings.</td>
</tr>
<tr>
<td>Assessment system</td>
<td><strong>ISCP</strong>: The system is available and documented online with access for trainees, educators and employers. This accessibility facilitates comparisons with previous assessments and helps ensure an appropriate rate of progression.</td>
</tr>
<tr>
<td>Content</td>
<td></td>
</tr>
<tr>
<td>General/Professional/Specialty specific</td>
<td><strong>AUS</strong>: Content covers both generic and specialty specific areas and contains a detailed description of the training structure including a description of the methods of learning and assessment. Clearly displaying the details of the developers and reviewers of each module contributes to keeping the content relevant to current practice. Learning outcomes are clearly benchmarked against levels of training.</td>
</tr>
<tr>
<td>Assessment system</td>
<td><strong>HK</strong>: A clear blueprint detailing the assessment process is freely available and easily accessible. Where minimum experience is required, such as operative volume, then these minimum numbers are made explicit and benchmarked to levels of training. While minimum numbers are stated, these do not directly indicate competence and assessment is still competency based.</td>
</tr>
<tr>
<td>Delivery</td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td><strong>CanMEDS</strong>: Clear description of how the curriculum can be used for Educators, Teachers, Trainees, Physicians, Researchers, Healthcare Professionals, Public Officials, Public and Patients. All programs are under the direction of a Canadian University Medical School. Clinical services that are used for teaching must be organized to promote their educational function. Programs lacking specific</td>
</tr>
<tr>
<td>Learning Model</td>
<td>resources can be accredited so long as formal arrangements have been made to send residents to other accredited residency programs to fulfil the deficient area.</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Learning Experiences</td>
<td><strong>AUS</strong>: Modular curriculum outline with clear objectives, suggested reading, description of learning opportunities and methods. Mandatory progression requirements include clinical experience, skills courses, in-training assessments, and examinations. A progressive increase in the amount of primary operator experience evidence in the logbook is expected. Research activity is also mandatory.</td>
</tr>
<tr>
<td>Assessment system</td>
<td><strong>CanMEDS</strong>: Strong emphasis on competency-based training as opposite to time based or volume based training. Consistent and logical structure and the overall concept of developing a surgeon who is more than just a clinician and has multiple roles.</td>
</tr>
<tr>
<td>Outcomes</td>
<td><strong>ISCP</strong>: Workplace based assessments are subject to reliability and validity measures. Assessments are standardised against predetermined and published criteria and stored in an online database allowing efficient delivery and monitoring of assessments and progress.</td>
</tr>
<tr>
<td>Supervision</td>
<td><strong>AUS</strong>: Trainees are each allocated a supervisor who is a Fellow in their chosen surgical specialty. Supervisors are nominated by the specialty training board and approved for a three-year term. RACS provide a number of training courses for supervisors.</td>
</tr>
<tr>
<td>Assessor</td>
<td><strong>HK</strong>: The trainee’s mentor conducts a face-to-face assessment as well as a logbook review every 3 months. In addition, twice a year the trainee submits their logbook and two mentor assessment forms for review by the specialty board.</td>
</tr>
<tr>
<td>Feedback</td>
<td><strong>AUS</strong>: A very transparent curriculum with clear and explicit details of the requirements for progression in terms of competency assessments, logbook numbers, courses and research. In addition, the timeline of what needs to be achieved at each stage of training and the processes for any grievance and appeals are also well described.</td>
</tr>
<tr>
<td>Documentation</td>
<td>13</td>
</tr>
<tr>
<td>Review and updating</td>
<td>14</td>
</tr>
<tr>
<td>Resources</td>
<td>15</td>
</tr>
<tr>
<td>Lay and patient involvement</td>
<td>16</td>
</tr>
<tr>
<td>Equality and diversity</td>
<td>17</td>
</tr>
</tbody>
</table>

**CanMEDS** = Canadian Physician Competency Framework, **ISCP** = Intercollegiate Surgical Curriculum Programme, **AUS** = Australian Surgical Curriculum, **HK** = Hong Kong Surgical Curriculum.
A limitation of this chapter is the use of GMC standards (GMC 2010) to compare the curricula. These are produced in the UK and thus the ISCP has been designed with these in mind. However no appropriate alternative criteria for assessing curricula could be identified. The purpose of using these standards was to produce objective and formative comparisons between the curricula in order to identify potential areas to develop rather than scoring the curricula to give an overall ranking. Using this rationale the GMC criteria were deemed appropriate.

While this chapter highlights global variability in surgical training programs e.g. in the structure, delivery and documentation of in-training assessments, objective comparisons of their respective surgical curricula revealed minimal differences. For instance the Australian, Hong Kong and ISCP curricula all meet standards 3 and 4 for content. However this does not reveal the differences between them in benchmarking to levels of training, which is heavily based on logbook numbers for the Australian and Hong Kong curricula but has much more emphasis on competency in the ISCP.

The challenge of how to accurately measure variability in surgical training programs remains. Current objective measures of comparison are based on quantity: how many years, how many hours, how many operations and how many patients seen? However a large scale study in the USA involving almost 4000 patients revealed that redesigning an inpatient medicine service led to training benefits despite a reduction in the numbers of patients seen (McMahon et al. 2010). Similarly when North American practice-ready surgeons were compared with European
practice-ready surgeons, their technical skills and cognitive knowledge were equivalent, despite the increased training hours of the North American surgeons (Schijven et al. 2010). Furthermore there is evidence that a focused competency based curriculum can accelerate surgical skills acquisition to the extent that trainees can move through the curriculum 12 months earlier than trainees following a regular curriculum (Ferguson et al. 2013; Sonnadara et al. 2011; Sonnadara et al. 2012; Sonnadara et al. 2014).

Quantity cannot be used as the sole marker of training quality. Following the paradigm shift from volume-based to competency-based training assessments of individual trainees, it seems perverse to rely on measures of quantity alone to measure the quality of surgical training programs. There is presently no assessment tool to evaluate the quality of surgical training programs that is analogous to those available for hospitals (DFRL 2012), universities (CUG) and research departments (REF). To improve surgical training there is first a need to objectively define high quality surgical training e.g. the ideal methods and amount of supervision and feedback, or the accessibility of technical skills simulators. Only once high quality training is defined can a valid assessment tool be developed to objectively evaluate the quality of surgical training. This will enable identification of potential target areas for development. An evidence-based approach has become the gold standard in patient care; it is essential to use an evidence-based approach to ensure training interventions can be objectively evaluated.

The reduction in work hours and its potentially detrimental effects on surgical training has been a major concern of the surgical profession (Antiel et al. 2013; Drolet et al. 2013; Morris-Stiff et
al. 2005). Reduced experience can certainly have an effect on surgical training. However, simply increasing time in training is not viable as the only solution. To use an analogy, the Formula 1 sporting industry has undergone a reduction in on-track testing times, which has parallels to the reduced clinical time facing modern surgical trainees. Despite this reduction in on-track testing, Formula 1 teams continually make advances such that a leading team can find themselves struggling in the midfield within the space of a season. The Formula 1 industry has embraced the use of simulation and maximise any data they receive from on-track testing for post-track analysis (BBC). Clearly surgical trainees cannot replace operative exposure by pouring over statistics created by a simulated procedure. However there is a need for a pragmatic approach to the real world situation of reduced clinical exposure. The current status presents an exciting opportunity to develop innovative methods to maximise the learning and feedback a trainee receives from every clinical encounter they are exposed to.

The question remains that if quantity is not the sole marker of quality then how can the quality of surgical training be demonstrated? In the USA, the American Board of Surgery In-Training Examination (ABSITE) is offered to General Surgery Residency programs annually to assess residents’ progress. The results are released to program directors and while not directly sent to individual residents, Program Directors will often inform residents of their results to provide feedback on their progress. The results may be used as a means to identify residents who would benefit from remediation. This could potentially be used as a part of a quality control exercise and is the closest to an assessment tool for surgical training quality that was identified in the surgical curricula examined. Perhaps a combination of something analogous to the
ABSITE along with Hong Kong’s flag criteria for minimum logbook numbers, could be used to develop a quality control system. This would take into account quantity while combining it with a measure of the quality of surgical training. The logical next steps would be to develop international expert consensus regarding what constitutes high quality training. Once high quality training can be defined, resources can be developed to measure the quality of training. Defining high quality training and the accumulation of data regarding the quality of training will enable the identification of examples of best practice. This information can be collated to improve national general surgical training curricula.

2.5: Conclusion

This chapter highlights variations between global General Surgical training programs, most notably in the length of training programs. Despite these and other variations such as the structure and delivery of the assessment system, when the standardised general surgical curricula of these nations are objectively compared, they appear remarkably similar. There is a need for an objective method of assessing surgical training quality that does not rely on quantity alone. An evidence-based approach has become the gold standard in patient care, it is now essential to invest resources into developing an evidence-based approach to ensure surgical training quality can be accurately evaluated to maintain and enhance standards.

In order to identify indicators of high quality surgical training, it is first important to identify the variations in surgical training programs. This chapter examined the differences between the
current surgical training programs globally using their respective surgical curricula. The next chapter will continue to examine variations in surgical training programs by gathering information from the trainees themselves. The focus will be on the UK and the USA training programs, as these appear to be the two that demonstrated the largest differences in this chapter’s evaluation of their respective curricula.
Chapter 3: A Comparative Study of Contrasting Surgical Residency Programs

The following chapter was published as:

3.1: Introduction

Work hours have been restricted to 80 per week in the USA and are further restricted to 48 per week in the UK as a result of the European Working Time Directive. While weekly working hours are longer, in the USA post-graduate training is completed in a five-year surgical residency program, compared to a minimum of ten years post-graduate training to reach consultant/attending level in the UK. Despite these differences, the common objective of both training programs is to produce trained surgeons capable of independent practice.

Chapter 2 highlighted a number of differences in the respective surgical training curricula of the USA and UK. With clear differences in the length and structure of training between the two countries, this chapter aims to further identify and evaluate differences between two surgical residency programs following their respective national training structure. While chapter 2 evaluated the training programs by looking at their documented surgical curricula, this chapter will evaluate the training programs from the perspective of the end users; the trainees.

3.2: Methods

Chapter 1 described how the paradigm shift from volume based training towards competency-based training has led to increased interest in the quality of surgical training. The UK’s Joint Committee on Surgical Training (JCST) has described quality indicators for surgical training(JCST 2012a). In the USA, similar indicators of quality can be inferred from the Accreditation Council for Graduate Medical Education (ACGME) Surgical Residency program requirements(ACGME...
and American Board of Surgery (ABS). Themes from the JCST quality indicators (JCST 2012a) were used to develop a questionnaire (see Appendix A). The questionnaire was piloted with three surgeons to ensure questions were not ambiguous and correctly focused on the desired themes. In the UK, medical training commences straight from high school and lasts for 5 years or 6 if an intercalated year of academic study to gain a Bachelors degree is undertaken. A 6-year surgical specialty registrar rotation begins from postgraduate year 5 following 2 years of Foundation training and 2 years of Core Surgical Training. Entry into each stage is competitive, unlike the run-through system in the USA. Independent practice as a Consultant Surgeon is only permitted upon successful completion of the surgical specialty registrar program. In the USA, following a 3 to 4 year pre-med course at a university, medical school is 4 years in duration. Medical students can go straight into their 5-year ‘run-through’ General Surgery residency program upon graduation and subsequently practice independently as an Attending surgeon.

Program directors at the University of Pennsylvania and the North and South West Thames General Surgery residency programs were invited to participate in the study. The Institutional Review Board granted an exemption for this study. Following approval, invitations to the study were distributed electronically by Program Directors via email to all general surgical residents in the University of Pennsylvania (n=64) and the North and South West Thames general surgical registrars in London (n=182).

The email invitation contained background information to the study. Trainees were informed that data collection was anonymous and participation was on a voluntary basis with no financial
rewards offered as a means of coercion. An electronic hyperlink link to the online questionnaire was provided and written consent was taken at the start of the questionnaire. In order to maximise responses, a reminder email was distributed 2 weeks after the initial invitation. All data was collected and stored in a secure online questionnaire system (http://www.surveymonkey.com). Data analysis was performed using STATA 12IC (College Station, TX) using non-parametric statistical comparisons. Where means and SDs are reported, Mann-Whitney U-test was employed. Where proportions of subjects are reported, a Fischer’s exact test or Chi² test was employed.

3.3: Results

Seventy-six trainees completed the online questionnaire: a response rate of thirty one percent. Twenty-nine responses were received from the University of Pennsylvania (UPenn) residents (45% response rate) while forty-seven responses were received in London (26% response rate). The demographics of the trainees are displayed in Table 3a. Trainees from London were older and in a higher postgraduate training year (PGY) than their UPenn counterparts.
**Table 3a: Demographic information, operative case volume and attending supervision level of London and UPenn residents**

<table>
<thead>
<tr>
<th>Demographics</th>
<th>London (Mean ± SD)</th>
<th>UPenn (Mean ± SD)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>34.4 ± 0.47</td>
<td>30.6 ± 0.60</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Postgraduate Training Year</td>
<td>9.19 ± 0.41</td>
<td>3.34 ± 0.39</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>n = 31</td>
<td>n = 22</td>
<td>0.455</td>
</tr>
<tr>
<td>Female</td>
<td>n = 16</td>
<td>n = 7</td>
<td>0.455</td>
</tr>
<tr>
<td><strong>Operative Case Volume</strong>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Surgeon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective Cases</td>
<td>2.83 ± 0.2</td>
<td>2.76 ± 0.3</td>
<td>0.580</td>
</tr>
<tr>
<td>Emergency Cases</td>
<td>2.04 ± 0.1</td>
<td>1.34 ± 0.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>First Assistant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective Cases</td>
<td>2.77 ± 0.2</td>
<td>2.82 ± 0.3</td>
<td>0.657</td>
</tr>
<tr>
<td>Emergency Cases</td>
<td>1.85 ± 0.1</td>
<td>1.54 ± 0.2</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>Supervision Level when Resident is Primary Operator†</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consultant/Attending in OR and scrubbed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective Cases</td>
<td>57.3 ± 4</td>
<td>82.9 ± 4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Emergency Cases</td>
<td>29.3 ± 2.9</td>
<td>89.8 ± 3.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Consultant/Attending in OR but not scrubbed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective Cases</td>
<td>21.9 ± 2.7</td>
<td>11.6 ± 3.6</td>
<td>0.003</td>
</tr>
<tr>
<td>Emergency Cases</td>
<td>10.6 ± 2</td>
<td>7.9 ± 3.6</td>
<td>0.045</td>
</tr>
<tr>
<td>Consultant/Attending in the hospital but not in OR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective Cases</td>
<td>4.5 ± 0.5</td>
<td>2.2 ± 0.4</td>
<td>0.001</td>
</tr>
<tr>
<td>Emergency Cases</td>
<td>23.9 ± 2.8</td>
<td>2.1 ± 0.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Consultant/Attending not in the hospital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective Cases</td>
<td>1.9 ± 0.3</td>
<td>1.2 ± 0.2</td>
<td>0.028</td>
</tr>
<tr>
<td>Emergency Cases</td>
<td>36.2 ± 3.7</td>
<td>0.2 ± 0.2</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* typical number of cases per week

† reported as a percentage of total emergency or elective procedures as recalled by trainees
All data presented is self-reported by the trainees. The volume of operating experience the trainees receive is displayed in Table 3a. No significant difference is demonstrated in the number of operative cases performed or assisted electively. Trainees in London performed significantly more emergency operative cases (mean 2.04 versus 1.34 cases per week p<0.001).

Table 3a also compares the level of supervision in the operating room (OR) for emergency & elective procedures. The level of supervision in the operating room was defined into 4 categories; (i) consultant/attending off-site but available, (ii) consultant/attending on-site but not in the OR, (iii) consultant/attending in the OR but not scrubbed and (iv) consultant/attending in the OR and scrubbed. When trainees operate electively, a consultant/attending is scrubbed in the OR 57% of the time in London versus 83% at UPenn (p<0.001). During emergency surgery, trainees operate without a consultant/attending in the OR for 60% of cases in London versus 2% at UPenn (p<0.001).

Table 3b compares activities in a typical day or night on-call as recalled by trainees; reported as a percentage of total shift. During a typical day on-call, trainees at UPenn spend significantly more time reviewing inpatients while London trainees spend significantly more time reviewing new admissions. During a night on-call, London trainees spend significantly more time operating in the OR and more time resting. Trainee resting hours count towards the duty hour limits in both the USA and the UK if the trainee is on-site.
Table 3b: Comparison of activities in a typical day or night on-call as recalled by London and UPenn residents reported as a percentage of total shift

<table>
<thead>
<tr>
<th>On-call activity</th>
<th>London (Mean ± SD)</th>
<th>UPenn (Mean ± SD)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rounding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day on-call</td>
<td>12.1 ± 1.3</td>
<td>9.1 ± 2.5</td>
<td>0.011</td>
</tr>
<tr>
<td>Night on-call</td>
<td>1.2 ± 0.4</td>
<td>4.4 ± 3.3</td>
<td>0.794</td>
</tr>
<tr>
<td>Reviewing new admissions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day on-call</td>
<td>33.2 ± 2.6</td>
<td>13.8 ± 1.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Night on-call</td>
<td>33.7 ± 2.6</td>
<td>29.3 ± 3.9</td>
<td>0.234</td>
</tr>
<tr>
<td>Reviewing inpatients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day on-call</td>
<td>13.4 ± 1.1</td>
<td>28.3 ± 3.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Night on-call</td>
<td>13.4 ± 1.3</td>
<td>39.1 ± 4.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Operating in the OR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day on-call</td>
<td>28 ± 2.2</td>
<td>34.1 ± 3.9</td>
<td>0.242</td>
</tr>
<tr>
<td>Night on-call</td>
<td>14.8 ± 1.3</td>
<td>8.5 ± 2.0</td>
<td>0.001</td>
</tr>
<tr>
<td>Elective commitments (e.g. clinic)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day on-call</td>
<td>6.4 ± 1.5</td>
<td>12.1 ± 3.4</td>
<td>0.010</td>
</tr>
<tr>
<td>Night on-call</td>
<td>0 ± 0</td>
<td>0.3 ± 0.3</td>
<td>N/A</td>
</tr>
<tr>
<td>Resting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day on-call</td>
<td>6.9 ± 2.2</td>
<td>2.7 ± 0.7</td>
<td>0.109</td>
</tr>
<tr>
<td>Night on-call</td>
<td>36.9 ± 3.2</td>
<td>18.3 ± 3.5</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Table 3c displays clinical activities. An operating session is defined as a half day operating list. UPenn trainees have significantly more operating sessions per week compared to London trainees (mean 5.0 versus 3.6; p = 0.005). In London 68% of trainees have regular gastrointestinal endoscopy sessions compared to 39% at UPenn (p = 0.036). Regular gastrointestinal endoscopy was defined as at least one endoscopy session per month. Trainees who indicated they were not in a gastrointestinal team were excluded for this analysis, leaving 38 London trainees and 18 UPenn trainees.

Table 3d demonstrates access and exposure to educational activities. Trainees at UPenn receive significantly more scheduled teaching (mean 2.3 hours per week versus 0.7 in the UK p<0.001). Almost all trainees at UPenn have access to laparoscopic box trainers, virtual reality laparoscopic simulators and endoscopic simulators while these are only available to a minority of London trainees.

Table 3e displays frequency of feedback from attendings as reported by trainees on a 5-point Likert scale (1-Never, 3-Sometimes, 5-Always). The amount of feedback received is largely similar however where significant differences are seen, the feedback is greater in London.
Table 3c: Clinical activities as reported by London and UPenn trainees

<table>
<thead>
<tr>
<th>Elective clinical activities</th>
<th>London (Mean ± SD)</th>
<th>UPenn (Mean ± SD)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportions of patients seen in clinic that are new i.e. not follow up (%)</td>
<td>61.7 ± 3.1</td>
<td>72.4 ± 3.6</td>
<td>0.025</td>
</tr>
<tr>
<td>Active role in preoperative timeout (% of operative cases)</td>
<td>89.4 ± 0.02</td>
<td>71.7 ± 0.06</td>
<td>0.063</td>
</tr>
<tr>
<td>Operating Room sessions (1/2 day) in a week</td>
<td>3.6 ± 0.2</td>
<td>5.0 ± 0.4</td>
<td>0.005</td>
</tr>
<tr>
<td>Clinics in a week</td>
<td>2.1 ± 0.1</td>
<td>1.9 ± 0.2</td>
<td>0.387</td>
</tr>
<tr>
<td>Opportunity for regular* GI Endoscopy (if in GI post) n = 26</td>
<td>(68%)</td>
<td>(39%)</td>
<td>0.036</td>
</tr>
</tbody>
</table>

*Regular defined as at least once monthly
### Table 3d: Access and exposure to educational activities as reported by London and UPenn trainees

<table>
<thead>
<tr>
<th>Teaching</th>
<th>London Mean ± SD</th>
<th>UPenn Mean ± SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching hours per week</td>
<td>0.7 ± 0.1</td>
<td>2.3 ± 0.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Protected teaching hours per week</td>
<td>0.5 ± 0.1</td>
<td>1.4 ± 0.3</td>
<td>0.019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Access to Simulators</th>
<th>London Number (%)</th>
<th>UPenn Number (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Reality Laparoscopy Simulators</td>
<td>9 (19)</td>
<td>25 (86)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Endoscopic Simulator</td>
<td>9 (19)</td>
<td>26 (90)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Laparoscopic Box Trainers</td>
<td>15 (32)</td>
<td>25 (86)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Human Patient Simulators</td>
<td>3 (6)</td>
<td>23 (79)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cadaveric Animals</td>
<td>5 (11)</td>
<td>2 (7)</td>
<td>0.702</td>
</tr>
<tr>
<td>Live, Anesthetized Animals</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0.382</td>
</tr>
<tr>
<td>Human Cadaver</td>
<td>0 (0)</td>
<td>5 (17)</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Table 3e: Feedback from attendings as reported by trainees on a 5-point Likert scale (1-Never, 3-Sometimes, 5-Always)

<table>
<thead>
<tr>
<th>Feedback</th>
<th>London Mean ± SD</th>
<th>UPenn Mean ± SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal feedback during an operation</td>
<td>3.57 ± 0.1</td>
<td>3.35 ± 0.2</td>
<td>0.236</td>
</tr>
<tr>
<td>Verbal feedback after an operation</td>
<td>3.19 ± 0.1</td>
<td>2.42 ± 0.2</td>
<td>0.002</td>
</tr>
<tr>
<td>Written feedback after an operation</td>
<td>1.91 ± 0.2</td>
<td>1.38 ± 0.2</td>
<td>0.020</td>
</tr>
<tr>
<td>Verbal feedback weekly</td>
<td>2.57 ± 0.2</td>
<td>2.27 ± 0.3</td>
<td>0.233</td>
</tr>
<tr>
<td>Written feedback weekly</td>
<td>1.72 ± 0.1</td>
<td>1.38 ± 0.2</td>
<td>0.044</td>
</tr>
<tr>
<td>Verbal feedback monthly</td>
<td>2.67 ± 0.2</td>
<td>2.54 ± 0.3</td>
<td>0.591</td>
</tr>
<tr>
<td>Written feedback monthly</td>
<td>1.89 ± 0.1</td>
<td>2.35 ± 0.3</td>
<td>0.518</td>
</tr>
<tr>
<td>Verbal feedback at end of rotation</td>
<td>4.09 ± 0.2</td>
<td>3.19 ± 0.3</td>
<td>0.003</td>
</tr>
<tr>
<td>Written feedback at end of rotation</td>
<td>3.55 ± 0.2</td>
<td>4.12 ± 0.2</td>
<td>0.149</td>
</tr>
</tbody>
</table>
3.4: Discussion

The large difference in time it takes to reach complete general surgery training may surprise surgeons on both sides of the Atlantic. The fact that trainees in London were older and in a higher training year is consistent with the longer training program and the later entry point into general surgical residency in the UK. In addition to the more stringent weekly work restrictions, which are 48 hours in London compared with 80 hours at Penn, other variations identified between training programs in this chapter may explain why the USA residency program is shorter and potentially more efficient.

The most striking variation is in the level of supervision in the OR. When trainees operate electively, a consultant/attending is scrubbed in the OR for the vast majority of cases at UPenn compared to just over half the cases in London. In fact during emergency surgery, trainees operate without a consultant in the OR for the majority of cases in London while attendings are present in the OR for almost all emergency cases at UPenn. Therefore trainees at UPenn receive significantly more supervised operating experience. This could be interpreted as evidence that training is more efficient with supervised operating. Alternatively, greater supervision may reflect comparatively less experience in the USA compared to the UK. Identifying the optimal method to recommend is therefore difficult.

Different payment mechanisms for surgeons in the USA and the UK may also play a role in the level of supervision. In the UK, consultants’ payment is unaffected by whether or not they are in the OR, while in the USA the attending needs to be present in order to bill for the case.
Consultants and trainees in London need to come to a mutual agreement regarding the complexity of cases the trainee can perform without direct supervision. Patient safety must be ensured and whether or not these differences in supervision affected the patient outcomes is beyond the scope of this study. However, there is evidence to suggest that patient safety standards are not compromised when competent residents operate independently (Borowski et al. 2008; K. K. Singh and Aitken 1999; P. Singh et al. 2014a). Nevertheless there is no formal credentialing system that recognises when a resident is competent to perform operations independently. There is also the potential that the reduced consultant presence in London is due to conflicts with alternative clinical and non-clinical commitments. The finding that UPenn trainees have limited experience of operating without an attending in the OR poses interesting questions about their ability and confidence to operate independently once they graduate from their residency program. These concerns are substantiated by a survey of North American Fellowship Program Directors in which only a third agreed that incoming fellows could independently perform 30 minutes of a major procedure without supervision (Mattar et al. 2013). This may in part explain the growing number of graduating residents in the USA pursuing further fellowship training which now exceeds 80% (Lewis and Klingensmith 2012).

Trainees at UPenn have significantly more operating sessions with on average at least one extra session per week. Despite the extra sessions, no significant differences were seen in the number of cases performed or assisted electively. Furthermore, trainees in London performed and assisted in significantly more emergency cases than their UPenn counterparts, although this may be indicative of the different structure and organisation of emergency surgery
between the two countries. In addition, trainees in London have better access to gastrointestinal endoscopy sessions compared to their UPenn counterparts. This may reflect a reduced importance attributed to endoscopic experience in the USA in an attempt to streamline residency training into a shorter period. However, the ACGME do stipulate a minimum of 85 endoscopy cases for all residents (ACGME 2010). It would be interesting to compare trainees on endoscopy-focused rotations in some USA residency programs, with the London trainees. At UPenn there is a 2-week endoscopy rotation for PGY2 trainees when they perform 50 colonoscopies. In the UK, trainees who wish to perform endoscopy seek Joint Advisory Group on Gastrointestinal Endoscopy (JAG) certification, although this is not mandatory. Competency based provisional certification criteria for colonoscopy (JAG), flexible sigmoidoscopy (JAG) or gastroscopy (JAG) are explicitly stated but require a minimum of 200 procedures for each type of endoscopy before assessment.

Differences in the organisation of emergency surgery may in part explain why during a typical on call, trainees at UPenn spend significantly more time reviewing inpatients while during a night on call, trainees in London spend more time operating in the OR and more time resting. In addition, the increased OR time overnight may be a consequence of London trainees’ greater experience of independent operating. The increased time resting overnight for trainees in London is somewhat surprising given trainees are restricted to 48 hours compared to 80 hours at UPenn. The 48 hour week in London is taken as an average number of hours per week over a 26 week reference period so in any one week, more than 48 hours can be worked so long as the average over 26 weeks is under 48 hours per week (BMA). Similarly, in the USA the 80 hour
weekly limit is averaged over 4 weeks and includes all on-site call activities and moonlighting (ACGME 2011a). This study did not specifically ask about protected sleep however there is evidence that protected sleep periods on call can be feasible and improve both duration of sleep and alertness following on calls (Volpp et al. 2012).

Trainees at UPenn enjoy greater educational opportunities than their London counterparts with more than three times as much scheduled teaching per week. In addition, nearly all UPenn trainees have access to simulation equipment compared to a minority in London. In fact, in the USA the ACGME Review Committee for Surgery state that ‘Simulation and skills facilities must be available for all program residents’ (ACGME 2013). Furthermore, proficiency on the Fundamentals of Laparoscopic Surgery (FLS) program (FLS) is a mandatory requirement for ABS General Surgery Certification (ABS). This emphasis on teaching for USA trainees may facilitate their ability to shorten the residency programme when compared to the UK. The amount of feedback trainees receive is largely similar; however where significant differences are seen, the feedback is greater in London. This may be a product of the successful introduction of the Intercollegiate Surgical Curriculum Programme (ISCP) and its increased focus on workplace-based assessments, which encourage regular structured feedback to trainees (Beard 2008).

While this chapter has further identified differences in the surgical training programs between UPenn and London, there are some limitations that should be considered. A greater response rate, in particular form London trainees, would improve the validity of the findings. The poor response rate from London is likely a reflection of the ‘survey fatigue’ that can develop amongst
trainees. No data was collected on the actual case mix and complexity of the surgeries performed. While data regarding the training post and subspecialty were collected, due to the small numbers involved, further subgroup analyses by subspecialty were not performed. Even allowing for the differences in the UK and USA training pathways, UK trainees were more senior than their UPenn counterparts. All data collected was self-reported and in addition, the responses were taken from a single institution in the UK and the USA. Nevertheless some of the differences observed in the current study are so striking that they are likely to represent major differences in training between the two nations.

This chapter relies on retrospective self-reporting of experience from surgical trainees and is therefore susceptible to recall bias. An example of this is the very small minority of UPenn trainees who reported not having access to simulators while the majority from the same program reported having access. Differences between the UK and USA healthcare systems particularly in the organisation of emergency surgery may account for some of the differences observed. The UK National Health Service that is free for all UK citizens, is heavily reliant upon trainees for service provision. In the UK trainees participate in an emergency service with an approximately one-in-six, to one-in-ten on call schedule, while in the USA trainees participate in the emergency on call service only when they are on an emergency surgery team. This chapter assumes that consultant/attending status is equivalent in both the UK and the USA. There is indeed some evidence to support this (Schijven et al. 2010). However, the UK training system combines general surgery with sub-specialist surgical training while in the USA sub-specialty training often extends beyond general surgery residency and into fellowship training.
3.5: Conclusion

This chapter has evaluated general surgery training in two programs following their respective national training structures in the USA and the UK. Variations in surgical training have been identified. The objective was to analyse these variations and identify potential improvements to surgical training. Few would disagree that the better access to simulators and greater amounts of teaching offered to trainees at UPenn are improvements that London training programs should strive towards. However, the differences in supervision arrangements and exposure to endoscopy training are more debatable. The heavily supervised training of UPenn residents with reduced focus on endoscopy training could be reflective of a more efficient training program focused on achieving core surgical competencies. However, the alternative view could be that trainees in London are trained to a greater level of independence than their UPenn counterparts. Indeed senior figures in the USA have highlighted that the evolution of residency training and the subsequent effects on resident independence and autonomy is an area that needs further analysis (Lewis and Klingensmith 2012).

There is a need for graded increase in independence as trainees gain more experience and demonstrate competency. The challenge for surgical training programs is to deliver this in a robustly assessed credentialing system that ensures patient safety is never compromised. The UK system attempts this with an informal trainer-trainee credentialing system, while the USA seek to maintain patient safety at the expense of independence. The optimal recommendation may be a compromise between the two with a USA style heavily supervised program in the early years, with a more objective and transparent system to grant independent operating
privileges to more senior residents. In order to provide guidance and clear recommendations to surgical training programs, there is a pressing need to develop consensus on the principles of high quality surgical training.
Chapter 4: Defining Surgical Training Quality:

Perceptions of the Profession

The following chapter was published as:

4.1: Introduction

The primary objective of a surgical training program is to produce competent surgeons capable of delivering high standards of patient care. It stands to reason that to maintain high standards of patient care, surgeons require high quality surgical training. A retrospective analysis of almost 5 million cases involving over 4000 clinicians from 107 US residency programs, demonstrated that patients treated by a graduate from the bottom quintile experienced an approximately 30% higher complication rate than those treated by graduates of a top quintile program (Asch et al. 2009). This provides empirical support that the quality of training surgeons received can be associated with patient outcomes.

Chapters 2 and 3 have demonstrated the variability in surgical training programs. Chapter 2 demonstrated variability in the documented curricula, while chapter 3 provides end-user information on the variability between the longest and shortest training programs identified in Chapter 2. The challenge arises in defining high quality surgical training. Without establishing this definition, training quality cannot be measured. In order to raise standards of training there must first be a robust method to measure surgical training quality.

This chapter explores the hypothesis that quality of surgical training is related to more than quantity or volume alone. The objective is to utilise user-informed opinion to define indicators of high quality training by balancing the views of those who deliver the training (consultants) with its recipients (trainees).
4.2: Methods

A qualitative, semi-structured interview approach was employed as it enabled a more detailed evaluation of subjects’ experiences than would be permitted by a questionnaire study. Consultant and trainee general surgeons were selected using purposive sampling techniques to gather a balanced perspective on the training relationship. This was a non-randomized technique to sample a range of subjects in terms of gender, trainee and consultant experience, multiple training programs and special interests etc. in order to give a realistic representation of surgeons in the UK. A full range of surgical experience was sampled to include surgical trainee representation of the full training range and consultants who had been trained both in the Calman training program (K. Calman 1993) and also more senior consultants who had trained before the Calman era. Surgeons were sampled from three different training regions within the UK. Consultants with a range of skills and interests were interviewed and included academic surgeons, community hospital surgeons and subspecialist fellowship training consultant surgeons. Patients are important stakeholders when considering surgical training. However, due to their limited experience in the delivery of training it was decided not to sample patients in this exploratory study. All subjects participated on a voluntary basis with no payments made to the subjects. Informed verbal consent was taken at the start of each interview and all results presented have been anonymised.

Each participant was asked a series of questions using an interview topic guide (Appendix B) to explore their views on indicators of quality in training, methods by which these could be robustly measured and potential interventions to improve the quality of surgical training. The
interview protocol was developed and piloted on 2 subjects before the interview topic guide was finalised. Interviews were conducted in private in a non-clinical setting. Each interview was audio-recorded and lasted 30 to 45 minutes. All interviews were subsequently transcribed verbatim.

To ensure accuracy, transcripts were cross-checked with the original recordings. In addition, a randomly selected sample of 25% of transcripts was ‘member-checked’ with the original participants to ensure accuracy. All interviews were analysed by the main researcher. Emergent themes were identified according to a coding framework. A second researcher analysed a randomly selected sample of 25% of the transcripts. Crosschecking ensured reliability of theme extraction. This is an approach that has been previously employed (Hull et al. 2013). Emergent themes were subsequently reviewed by three researchers. Coding was performed by discussion, with disagreement resolved via consensus. Sampling was due to be ceased when thematic saturation was achieved.

4.3: Results

4.3.1: Subject Demographics

A total of 20 interviews were conducted with balanced numbers of trainees (n = 10; 7 Male: 3 Female) to consultants (n=10; 9 Male: 1 Female). Thematic saturation was achieved in fewer than 20 interviews however further interviews were conducted in order to achieve the balance of demographics that had been intended with the purposive sampling technique. Trainees’
experience ranged from junior trainees in their basic surgical training rotations, to final year
trainees focusing on their subspecialty interest (median = Postgraduate year (PGY) 7, range:
PGY 4-10). UK surgical training begins in PGY 3 with basic surgical training followed by
competitive entry at PGY 5 into a 6-year specialist general surgery training program as a higher
surgical trainee. The sample of trainees included 5 junior trainees (PGY 4 – PGY 6) and 5 senior
trainees (PGY 7 – PGY 10). No PGY 3s were sampled as they had just begun their surgical
training so they would have very limited experiences of the specialty training. Consultant
surgeons’ experience ranged from those recently qualified under the Calman training
programme(K. Calman 1993) to senior consultants and heads of departments (median = 4.5
years; range 1.5 – 16 years of consultant experience).

4.3.2: Indicators of high quality surgical training

Interviews revealed unanimous agreement on the existence of variability in the quality of
surgical training. Explanations cited for this variability included; exposure to cases, the
relationship between trainee and trainer, time constraints and whether or not the training
rotation was appropriate to the level of a subject’s training. The conflict between service
 provision and training was often cited, for example one consultant mentioned that,

‘there is also a massive amount of pressure on the consultants to get through a certain amount
of workload, waiting lists, access to theatre times and a few other things that put pressures on
the amount of time that they can actually give a trainee’.
Another reason cited was the consultants’ skill as a trainer;

‘it’s not just about being supervised in your training, but about being supervised by somebody who’s trained to supervise you and teach you’.

Further interrogation of the causes for this variability identified a number of indicators of quality in surgical training. The top 10 indicators are summarised in Table 4a. The interaction between trainer and trainee was identified as the most important indicator of quality, closely followed by the clinical exposure achievable in a rotation, both in terms of volume and diversity of pathology. In addition to traditional aspects in terms of feedback and the teaching programme, organisational aspects of a department and its effects on team structure were also identified as indicators of quality;

‘feeling part of the team, feeling like it makes a difference if I’m there every day. Actually being able to go into clinics and theatres, as I say, on a regular basis, to build up that rapport with your trainer’.

There was generally good agreement between the views of trainees and consultants. However, the majority of consultants identified a stepwise approach to training as an important marker of quality compared with only a minority of trainees. A stepwise approach to training is the ability to break down training milestones into smaller steps, sometimes with a graded increase in difficulty, that can be taught individually and then combined to demonstrate a graded progression in the trainee’s ability. Trainees were unanimous in their support of supervision as
a marker of quality although both trainees and consultants acknowledged the need for a balance between supervision and independence.
### Table 4a: Top 10 indicators of quality in training

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Number of Trainees (n = 10)</th>
<th>Number of Consultants (n = 10)</th>
<th>Example Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Trainer – trainee interaction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Relationship between trainee and trainer</td>
<td>9</td>
<td>10</td>
<td>‘Even, actually, a very busy weekend on call can be enough to develop a relationship with a team that you’ve not met before and if you end up doing 12 operations with those people over a long weekend on call, for example, that can be enough to build up a useful training relationship’</td>
</tr>
<tr>
<td>• Willingness of consultant to allow cases to be performed by the trainee</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>• Consultant engaged with the training process/curriculum</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>• Willingness of consultant to teach</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>• Approachability</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2. Exposure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Case Mix</td>
<td>10</td>
<td>8</td>
<td>‘There must be workload. It doesn’t have to be excessive. In fact, excessive workload is probably a mistake because people get too busy and then it doesn’t provide a training opportunity’</td>
</tr>
<tr>
<td>• Logbook of previous trainees</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>• Volume of cases</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>• Balance of emergency v elective workload</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3. Supervision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Level of supervision</td>
<td>10</td>
<td>7</td>
<td>‘Yes, it’s supervised independence which sounds a little bit nonsensical. I’m around but I am less visible but I am aware of what’s going on. We record everything that happens in theatre through the laparoscope. We don’t get the external view but we record everything that happens in theatre’</td>
</tr>
<tr>
<td>• Balance between supervision &amp; independence</td>
<td>8</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>• Stepwise increase in independence</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>4. Good feedback from</td>
<td>8</td>
<td>9</td>
<td>‘We always measure patient experience and</td>
</tr>
<tr>
<td>previous trainees</td>
<td>you always ask the patient what their experience was like and I don’t think it’s unreasonable to ask a trainee what their experience was like’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. <strong>Feedback</strong></td>
<td>‘Two minutes at the end of the list or at the end of the case just them summarizing to you your good bits and your bad bits and where to aim for’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Informal</td>
<td>7 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Regular &amp; constructive</td>
<td>5 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• During or after every case</td>
<td>5 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• During and after every operating list</td>
<td>5 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Formal</td>
<td>4 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. <strong>Organisational</strong></td>
<td>‘It was not uncommon for the consultant to have the same registrar, the same SHO and the same houseman for a six month period and the whole working week would involve the same people and the on calls as well. Now here, for example…..you never quite know who is going to be around at any given day. It is very difficult to allot the same level of trust to a registrar that you’re not familiar with working with’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Timetable Structure</td>
<td>8 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Continuity with the same consultant</td>
<td>7 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Balance between time constraints and training</td>
<td>8 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Good firm/team structure</td>
<td>5 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• On-call and departmental structure</td>
<td>4 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Trainer : Trainee ratio/competition for training</td>
<td>6 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. <strong>Evidence of progress during the job</strong></td>
<td>‘we’ve looked at the ways that young surgeons are gaining proficiency and we see them reaching a stage of independence in laparoscopic colorectal surgery that might be a little bit quicker than other people’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. <strong>Personalised training</strong></td>
<td>‘I think it’s where your job and your trainer, where their expectations are at the right level’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Training placement</td>
<td>4 8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
appropriate to level of training

- Training to meet educational needs

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

of your level of ability so if you meet each other at the right stage in life, it’ll be fine. For example, I’ve worked in places where things are below my level and I’ve just done things that I did two years earlier and I really didn’t learn a lot more. Equally, I’ve been in a very junior position where I’ve worked in units where things have been way above my level; I’ve not been able to get involved too much’

9. Teaching Programme

- Access to simulators
- Structured teaching programme
- Ability to attend courses

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

‘endoscopy and laparoscopic work really does lend itself to potentially training and simulation I think better than anything else and I think if they get you to a skillset where you’re comfortable that then allows you to actually operate in the real world then that’s a brilliant training’

10. Stepwise approach to training

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

‘My best training job was by a laparoscopic colorectal surgeon who broke down very complex procedures at a time when they still were regarded as extremely complex into a series of steps and it was taught to us almost like a sort of cooking recipe’
In order to assess the effect of experience on opinions, trainees were grouped into junior trainees (PGY 3 – 6) and senior trainees (PGY 7 – 10), giving an even distribution of 5 in each group. Similarly consultants were grouped as junior consultants (less than 5 years since appointment) and senior consultants (5 or more years since appointment), which again provided an even distribution of 5 in each group. On the whole opinions were fairly similar. Senior trainees were more likely to mention the on-call and departmental structure as an indicator of quality (0 junior trainees, 4 senior trainees) with a particular preference for 24-hour shifts. All senior trainees identified evidence of progress during the rotation as an indicator of quality while no junior trainees specifically mentioned this. Junior consultants more frequently identified access to simulators (5 junior consultants, 1 senior consultant) a structured teaching programme (4 junior consultants, 0 senior consultants) and the ability to attend courses (4 junior consultants, 0 senior consultants) as indicators of high quality training.

4.3.3: Metrics to measure quality of surgical training

The top 10 metrics identified for measuring the quality of training are shown in Table 4b. A multifactorial approach emerged with subjective factors, such as trainee evaluations of consultants, combined with more quantifiable objective parameters such as logbook volume. In addition to feedback from trainees, consultant feedback was felt to be essential by the majority of subjects (60%; 7 trainees, 5 consultants). While some metrics identified already exist, other metrics were described that would require modifications to current data capture systems such as the operative logbook.
### Table 4b: Top 10 Methods of Measuring Quality in training

<table>
<thead>
<tr>
<th>Metric to Measure</th>
<th>Number of Trainees (n = 10)</th>
<th>Number of Consultants (n = 10)</th>
<th>Example Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Trainee feedback</td>
<td>10</td>
<td>10</td>
<td>‘whether we’d like it if we didn’t get a good rating - there’s obviously a need but I think it would be a good stimulus to improve’</td>
</tr>
<tr>
<td>* Use aggregate measures</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2. Trainee logbook for case mix and case load</td>
<td>10</td>
<td>10</td>
<td>‘it comes down to the kind of ratio. So the number of cases that are done when the trainee is present and how many of those cases that trainee has actually done’</td>
</tr>
<tr>
<td>* Use aggregate measures</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. Formative evaluation of consultants by trainees</td>
<td>6</td>
<td>9</td>
<td>‘We often talk to our trainees at any periods along the way and they should be able to assess us as we should be able to assess them. I think that’s very important’</td>
</tr>
<tr>
<td>4. Effect of on-calls and timetable on continuity</td>
<td>6</td>
<td>8</td>
<td>‘I think the problem is that the rota is everything and the rota comes first and then the training is an incidental’</td>
</tr>
<tr>
<td>5. Consultant feedback</td>
<td>7</td>
<td>5</td>
<td>‘very much, yes. Trainers will often voice concerns about training and I think that’s very valid’</td>
</tr>
<tr>
<td>6. Modify the logbook to allow greater detail</td>
<td>3</td>
<td>6</td>
<td>‘we look at the number of modules that a surgeon does or achieves and as they’re getting more and more then we see globally.......you actually measure their progression through the modules as well.......Initially that was done in an informal way but now we get them to record events that they do and it allows us to target areas that are deficient’</td>
</tr>
<tr>
<td>7. Objective numbers from</td>
<td>4</td>
<td>4</td>
<td>‘I think for things like cancer, for example, you’</td>
</tr>
</tbody>
</table>
the hospitals
e.g. number of cases,
number of training lists
could do it. This data is collected through the cancer databases certainly...I think it is quite important for senior GI trainees to have their fair share of the cancer work’

<table>
<thead>
<tr>
<th>8. Evidence of trainee improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Evaluation of trainee by attending</td>
</tr>
<tr>
<td>• Assessment of trainee</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>‘And then you can say, “Okay, at this center, 90% of our trainees achieved a 20% improvement in their core knowledge and also their practical skills.” And then you can rank all... the centers, and say, “This center has a 90-whatever percent success rate in producing a 10% improvement.’”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. Consultant logbooks</th>
</tr>
</thead>
<tbody>
<tr>
<td>What proportion of cases do they let their trainees do</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>‘you could look at the departmental logbook...For instance, they’ve looked at trauma centres and looking at the number of trauma cases and said how many of these trauma cases are actually done by the trainee’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. Benchmark minimum numbers of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
</tr>
<tr>
<td>‘there should be national standards for training posts. If you’re going to be a colorectal training post, you need your trainee to have achieved so many laparoscopic resections in a year or something like that and set a national standard.....’</td>
</tr>
</tbody>
</table>
Almost half the subjects felt more detail in the logbook would be beneficial; one consultant explained

‘if there was some way of breaking down every operation down into a checklist and different components and trainees showed what bits they can do and what bits they have done that would tell me more about the unit’.

For instance there can be a marked improvement when assisting in complex operations from being a passive assistant, to competently performing significant steps of the operation.

While both trainees and consultants mentioned assessment and evaluation of trainees, consultants were more explicit in stating that it is the improvement in trainee performance over time that would be a better marker than outright performance. Trainee feedback and logbook analysis for both volume and case mix, topped the metrics to measure training quality with unanimous support from both trainees and consultants. Subjects voiced concerns that current trainee feedback is likely to be inadequate as trainees can be reluctant to raise any training concerns to their immediate seniors. In order to alleviate this difficulty, aggregate measures were specifically mentioned by a number of subjects to avoid identification of individuals’ feedback.

There was unanimous agreement that trainees should be involved in performing any proposed evaluation of training quality and the majority (60%; 7 trainees, 5 consultants) felt consultants
should also be involved. Only 6 subjects (30%) felt independent assessment was needed although this was largely in acknowledgment of the need for pragmatism in that independent assessment would be costly and perhaps unfeasible for all training rotations. Only a minority of subjects identified benchmarking minimum numbers of cases as a measure of training quality;

‘you could put a minimum on that and say you should be doing a minimum number of laparotomies or laparoscopic procedures in a given year but I think again that’s quite difficult.... it would probably need to be individualized, personalised’.

The overwhelming majority of subjects felt the results of training quality evaluation should be made transparent to all (85%; 7 trainees, 10 consultants). Subjects felt the ability to measure quality of training would have an overall positive effect on training as transparency would encourage competition (70%; 6 trainees, 8 consultants) and create accountability for hospitals and consultants (70%; 7 trainees, 7 consultants). In addition, a number of subjects felt the information gleaned from measuring training quality would enable a more personalised training programme to be provided by facilitating the matching of trainees to the most appropriate training rotations according to their level of experience (45%; 3 trainees, 6 consultants).

4.3.4: Interventions to improve training quality

The most frequently suggested intervention was to encourage trainee feedback (at the end of their rotation) about the exposure an incoming trainee can expect to gain from a particular rotation (70%; 6 trainees, 8 consultants). There was also support for improvements to the team
structure and to increase continuity of exposure to the same consultant (65%; 5 trainees, 8 consultants);

‘If you’re in a hospital where there are only six general surgical consultants, you’re going to get a decent relationship with probably all six over a year. If you’re in a place where it’s got 12 to 15, you’re probably not and so therefore if you’re at a bigger center then maybe you should be allocated three or four consultants on a regular basis that you’re on call with rather than being spread about across the whole 15 each year’.

Subjects identified the concept of personalised training through more detailed placement of trainees to training rotations that are suited to their level of training (60%; 6 trainees, 6 consultants);

‘If you want to be a colorectal trainee and you go to a colorectal center as a Year 5 that isn’t doing any laparoscopic surgery, it’s probably not going to be what you want to do. If you go to that center and you’re Year 1 and they’re doing a lot of open colorectal surgery, it might actually be ideal for you’.

Eleven subjects (55%; 6 trainees, 5 consultants) felt access to simulation based skills training would improve the quality of training and in particular there was enthusiasm for supervised technical skills training.
'There are various benefits to having a simulator available at any time for a trainee but I think there should be sessions where a trainer is available, can supervise the performance and could assist in improving the performance of the trainee’

Other potential interventions identified included providing feedback to the consultants on how well they train and what they can do to improve (50%; 2 trainees, 8 consultants); modifying the logbook to include completion of parts of operations (40%; 3 trainees, 5 consultants); implementation of a structured teaching programme to include practical skills and lectures (40%; 4 trainees, 4 consultants); and training operating lists (35%; 5 trainees, 5 consultants) with increased trainee input into the case selection for each list. Incentivising high quality training was suggested as an intervention in a number of different ways. Firstly through reward; financial or otherwise for consultants (35%; 4 trainees, 3 consultants) or by compensating hospitals that are prepared to train trainees (30%; 4 trainees, 2 consultants). Alternatively through punitive measures such as removing trainees from low quality training rotations (35%; 3 trainees, 4 consultants).

4.3.5: Should all hospitals and consultants train?

Whether all hospitals and consultants have a duty to train their juniors was a contentious subject with a majority rejecting this notion (65%; 7 trainees and 6 consultants). However, there were some polarised views on this subject as demonstrated by one consultant who stated;
'We have to get away from this idea that as a doctor or as a consultant, we’re all the same.

We’re not all there to be management. We’re not all there to be educators’.

Another consultant supported this view that being a trainer was not for everybody;

‘You’ve got to have a desire to do it, you’ve got to have a skill at doing it and you’ve got to have the workload to do it and also the management that supports that’.

However, another more senior consultant was equally firm in his contrasting view that training was a duty incumbent upon all clinicians;

‘I think it’s part of our Hippocratic Oath, is to train the next generation of doctors. And I think when we become doctors and surgeons, we take that as part and parcel of training the next generation. And I think forgetting to train your junior is forgetting your Hippocratic Oath’

4.3.6: Whose responsibility is it to ensure the quality of training?

Subjects were asked whose responsibility it was to ensure the quality of surgical training was meeting the expected standards. The results revealed little consensus with 13 different bodies mentioned including the government, Higher Surgical Training Committee and individual hospital chief executives. The most popular answers were the Deanery (35%; 3 trainees and 4 consultants), the Royal College of Surgeons (30%; 5 trainees and 1 consultant), individual
consultants (30%; 2 trainees and 4 consultants) and the trainees themselves (25%; 4 trainees and 1 consultant). One trainee stated;

‘Morally, it’s the trainees’ responsibility, I’ve got responsibility to myself and to my potential patients to train myself, and I think if there’s any deficiencies, that’s up to me.’

4.4: Discussion

The objective of surgical training is to produce competent surgeons capable of maintaining high standards of patient care. There is evidence to suggest the quality of training received can be associated with patient outcomes (Asch et al. 2009). This chapter sought to define high quality surgical training and subsequently to explore methods by which it can be measured. Robust tools to evaluate surgical training quality will enable surgeons and hospitals to demonstrate the quality of surgical training they can provide. Once the definition of quality was explored, ideas for focused training interventions to optimise quality were discussed.

Thematic saturation was achieved and revealed a number of common themes when defining high quality surgical training. Trainees need a progressive increase in independence, not just in the operating room but also with non-technical decision-making skills outside the operating room. One notable difference in responses between consultants and trainees was that the majority of consultants identified a stepwise approach to training as an important marker of quality compared with only a minority of trainees. This is perhaps indicative of the greater experience of the consultants giving them a more holistic view of what is required to produce
competent surgeons. Trainees tend to focus more on whole cases and perhaps do not appreciate how much is learnt by performing parts of operations. This is potentially due to the fact that presently, numbers of cases performed is one of the few objective ways in which trainees can compare themselves against their peers and also with themselves when monitoring their own development.

On measuring training quality, a novel concept proposed by some subjects was to evaluate consultant logbooks and in particular the proportions of operations they permit their trainees to perform. This is an innovative way to develop a metric that combines trainer involvement while taking into account the throughput of a unit, thus accounting for factors that are out of the consultant’s control. However, it would potentially penalise consultants that intentionally work through a number of cases quickly in order to allow extra time to take trainees through a smaller number of cases in greater detail. Perhaps surprisingly, despite the recent introduction of guidelines explicitly stating minimum numbers of operative cases by the JCST (JCST 2012b), only a minority of subjects identified benchmarking of minimum numbers of operative cases as a metric to measure training quality.

When discussing training interventions, more trainees identified adding time to training lists compared with consultants who were more interested in improving the continuity of exposure between a trainee and consultant. Consultants expressed concerns about providing training to trainees that they do not see on a regular basis. Consultants were very enthusiastic about receiving feedback on their training quality while trainees did not focus so much on this. A
number of explanations for this could be postulated, not least because consultants are naturally likely to focus more on their own involvement. Another explanation is that trainees assume consultants have insight into their training ability; i.e. a poor trainer knows they are poor at training. In addition there was an impression elicited from trainees that poor trainers did not have an interest in training. In contrast, a common theme to emerge from interviews with consultants was how little insight they actually had into their training ability and how they would welcome constructive feedback in order to improve the training they could offer;

‘I have absolutely no idea whether my training is good or bad and whether I’m universally good or bad.......and the same goes for all my colleagues. So none of us knows which one of us is a good trainer and which one is a bad trainer and it irritates me as somebody who is in medical politics as well that there is this assumption that because you’re a consultant you’re a trainer and you’re bound to be good. It’s a kind of arrogance....’.

‘Train the trainers’ courses may prove a useful intervention in these instances where consultants found to be poor trainers require educating on how to become a better trainer.

This study sampled the views and opinions of 20 surgeons currently practising in the UK healthcare system. Subjects were required to make immediate responses to questions rather than allowing a considered response after a period of reflection. The generalisability of these results could be questioned. However, both consultants and trainees with a wide range of experience were purposively sampled to minimise selection bias. In 2011, 26% of surgical
trainees and 8.7% of consultant surgeons in England were women (RCSEng). This study sampled 3 female trainees (30%) and 1 female consultant surgeon (10%) so it provides a realistic representation of the proportions of women in surgery. The range of experience covered those who had trained in the older traditional training systems with long hours and an abundance of operative cases as well as the current cohort of trainees more accustomed to restricted training hours but arguably more structured training. The fact the study reached saturation of emergent themes in the main subject areas is a demonstration of the success of the purposive sampling technique. This study did not sample the opinions of patients. However, while patients are important stakeholders regarding any proposed reforms to surgical training, it was felt that their lack of in depth knowledge about how training is delivered would preclude them from being able to contribute significantly to this exploratory study.

The results presented demonstrate the potential implications of measuring training quality both in surgery and potentially all branches of medicine. Financial incentives for healthcare service delivery have attracted much attention with reports that they can also be related to significant improvements in patient outcomes (Sutton et al. 2012). With proposed reforms to educational commissioning in the UK, hospitals need to demonstrate evidence of high quality surgical training in order to receive surgical trainees. Trainees play a crucial role in the running of general surgical units and in particular the out of hours general surgical cover. The loss of surgical trainees from a unit and their associated training budget would have a significant impact both on the hospitals concerned and also to the affected regions’ trainee experience. There are however other incentives for a healthcare institution, not least their reputation
amongst other institutions. In the UK a unit needs to demonstrate excellence in education, research and patient care to earn Academic Health Science Centre (AHSC) status (DoH 2008b; Dzau et al. 2010). Institutions have undoubtedly concentrated their resources into improving research and clinical outcomes in order to perform well on the nationally reported Research Assessment Exercise (RAE 2008) and Dr Foster league tables (DFRL 2011). In order to redress the balance of resources back towards surgical training outcomes, a tool to robustly and objectively assess surgical training quality needs to be developed to provide a similar incentive to improve surgical training programs.

In order to build on the results of this chapter and provide evidence-based recommendations that can be applied globally, international expert consensus on the indicators of high quality surgical training is necessary. These findings will inform and facilitate the next goal, which is to develop an assessment tool capable of objectively evaluating the quality of a surgical training program. Furthermore, identifying indicators of high quality surgical training will subsequently permit focused training interventions to be developed that can maximise the quality of surgical training. In the current climate of reduced clinical exposure and the drive for increased efficiency across healthcare, maximising the quality of training is a crucial endeavour. Defining who is responsible for this is an interesting and perhaps surprisingly, a challenging question. This chapter revealed little consensus with 13 different bodies mentioned including the deanery, government, trainees and individual hospital chief executives.
4.5: Conclusion

Surgical training quality needs to be robustly assessed and results should be made freely available. Surgeons believe creating a faculty of nominated trainers alongside transparency in training outcomes will create competition and raise standards of surgical education. The results presented provide a platform to develop international consensus on the optimum methods to evaluate surgical training quality.
Chapter 5: A Global Delphi Consensus Study on Defining & Measuring Quality in Surgical Training

The following chapter was published as:

5.1: Introduction

Traditional methods employed to evaluate surgical training quality use surrogate markers of quality. For example, number of cases or number of hours worked\cite{morris2005}. However there is global pressure to reduce training hours. Work hour restrictions, introduced in the UK and the USA, that further reduce trainee exposure and caseload\cite{drake2013}, have created much discussion\cite{drolet2012, rosenbaum2012}. Chapter 1 describes the UK’s Temple report which was commissioned to evaluate the effects of the European Working Time Directive (EWTD)\cite{temple2010}. It concluded that ‘extending the hours worked or the length of training programmes are not sustainable solutions’. Similarly, in the USA the Accreditation Council for Graduate Medical Education (ACGME) set up a Task Force in 2009 to comprehensively review the impact of the implementation of the 2003 ACGME implemented common duty hour standards. Following an international duty hour symposium, 3 commissioned literature reviews and a national duty hours congress, the Task Force recommended more restrictive duty hour limits for first year residents while adding flexibility for senior residents\cite{acgme2011b}.

Chapter 1 describes how proposed reforms to educational commissioning in the UK place greater responsibility on hospitals to demonstrate high quality surgical training in order to be assigned trainees and their associated training budgets. In the USA, the Residency Review committees of the ACGME set standards for residency programs. Non-compliance results in probation and ultimate failure to meet standards can result in withdrawn accreditation of the program. The ACGME Task Force examining the 2003 Duty Hours reforms ‘affirmed that the
standards would need to go far beyond limits on resident hours to promote high-quality education’ (ACGME 2011b). There is currently no accepted definition of what constitutes high quality surgical training. To enable hospitals and surgical training programs to provide evidence of training quality, it is first necessary to define the elements of high quality surgical training and metrics for measuring them.

In chapter 4 an interview study sampling both trainees and trainers was conducted to identify indicators of high quality surgical training both inside and outside the operating room (P. Singh et al. 2014b). The objective of this chapter is to use the information gleaned from chapter 4 to develop consensus among international surgical education experts on the indicators of high quality in surgical training and how they can be measured. This information will subsequently be used to develop a tool to evaluate surgical training quality.

5.2: Methods

5.2.1: Study Design

A Modified Delphi Methodology was employed in order to gain expert consensus on the indicators of high quality surgical training and how to measure them. The Delphi technique is an iterative process that employs a number of anonymous voting rounds in which the participants never meet. This classical Delphi methodology removes the geographical and financial constraints of requiring experts to meet in person (Fink et al. 1984). The anonymity also reduces the disproportionate impact that a single influential participant can have on the rest of the
participants (Fink et al. 1984; Graham et al. 2003). A criticism of the fact that the participants do not meet is that no discussion takes place (Graham et al. 2003). It is for this reason that a modified Delphi methodology was employed, which allowed for discussion to take place at a workshop during an international surgical education conference. No voting took place during the workshop to ensure anonymity was maintained during the voting process.

5.2.2: Identification of international panel of experts

Experts in the field of surgical education were identified from committees of international surgical education associations which included representation from the International Conference on Surgical Education and Training, the Association for Surgical Education, the World Congress on Surgical Training, the American College of Surgeons Division of Education, the Society of American Gastrointestinal and Endoscopic Surgeons and academics who have published in the field of surgical education. In addition, senior members of National Surgical Colleges and Surgical Trainee association representatives were identified and included representation from the Royal College of Surgeons of England, Ireland and Edinburgh, the American College of Surgeons, The Royal Australasian College of Surgeons, The College of Surgeons of Hong Kong, the College of Surgeons Singapore, the Canadian Association of General Surgeons and the Association of Surgeons in Training. A purposive sampling technique was employed to obtain representation from a truly international panel of experts. A total of eighty-three experts in surgical education from 13 countries were identified.
5.2.3: Identification of themes for the Delphi survey

In chapter 4, twenty semi-structured interviews were conducted with General Surgery consultants (10) and General Surgery trainees (10) in the UK (P. Singh et al. 2014b). Seventy statements regarding indicators of training quality and methods by which these could be robustly measured were developed based on themes identified from the semi-structured interviews of consultant and trainee general surgeons.

5.2.4: Administration of the survey

The survey was administered online (www.surveymonkey.com). Personalised invitations were sent to the expert panel via electronic mail with a hyperlink to the online questionnaire. Participation in the study was voluntary with no financial or other remuneration offered. Consent was taken at the start of the survey. Experts were asked to rank each statement on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). A free text comment box was also available at the end of each statement. Two reminder emails at two-week intervals were sent after each round of voting. Interim results of the first round of voting were presented in a Workshop at the International Conference on Surgical Education and Training, Ottawa, Canada, 2012. The results presentation focused on the borderline statements that were close to the pre-determined acceptance criteria. This was intended to encourage open discussion and debate prior to any further rounds of voting. For subjects unable to attend this Workshop, a summary of results and feedback from the first round of voting and discussion was sent to all participants electronically before commencement of a subsequent round of voting. This included the mean scores for each statement. All results and feedback were anonymised so no
individual or institution could be identified. The online survey was repeated until consensus had been achieved according to the predefined criteria.

5.2.5: Statistics and Definition of Consensus

Cronbach’s α was chosen as the statistical test to quantitatively evaluate consensus amongst the international expert panel. A Cronbach’s α > 0.90 has been suggested as necessary for clinical diagnostic scales (Bland and Altman 1997; Shrout and Fleiss 1979). Previous studies have set a Cronbach’s α of ≥ 0.80 as representative of an acceptable measure of internal reliability (Nagpal et al. 2010; Zevin et al. 2013). Best practice guidelines suggest that criteria for consensus should be defined in advance (Fink et al. 1984). Therefore consensus was predefined as Cronbach’s α ≥ 0.80.

5.2.6: Development of the S-QAT

Statements that ≥80% of experts ranked as ≥4 (agree or strongly agree) were used as themes to develop the Surgical training - Quality Assessment Tool (S-QAT). A Likert scale was used to assign descriptive ratings for each of the indicators of surgical training quality. The use of a Likert scale was chosen due to its ability to bring a degree of objectivity to sometimes subjective measures. Furthermore, its frequent use in some of the most widely used assessment tools within the field of surgical education (Martin et al. 1997; Vassiliou et al. 2005), was seen as an advantage as it would be familiar to most end-users of the tool.
5.3: Results

A total of 53 (64%) experts from 11 countries participated in the two rounds of the Delphi survey. Figure 5a displays the international representation for the surgical education expert Delphi panel. In the first round 47 (57%) experts from 10 countries responded and completed the survey. Thirty-three of the 70 statements achieved consensus with a Cronbach’s α of 0.904 when means were used to replace missing values, 0.904 when modes were used for missing values and 0.906 if 3 (neutral) was used for missing values. Forty-three people attended the Workshop at the International Conference on Surgical Education and Training, Ottawa, Canada, 2012. In the second round of voting, 35 (42%) experts from 10 countries responded and completed the survey. Six experts who responded were new experts that had not responded in the first round of voting, thus giving a total of 53 experts (47 from first round plus 6 new experts in the second round) from 11 countries who participated in the study. In the second round 35 of the 70 statements achieved consensus with a Cronbach’s α of 0.930 when means were used for missing values, 0.930 when modes were used for missing values and 0.930 when 3 (neutral) was used for missing values. The statements that reached consensus according to the predefined criteria are shown in Table 5a. The statements that reached consensus were used to develop the S-QAT tool (see Figure 5b). The statements that did not reach consensus can be found in Appendix C.
Figure 5a: International Representation of Surgical Education Experts
Table 5a: Statements reaching pre-defined consensus criteria for acceptance

<table>
<thead>
<tr>
<th>STEM</th>
<th>STATEMENT</th>
<th>MEAN</th>
<th>SD</th>
<th>MISSING DATA %</th>
<th>% RATING 4 or 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measuring the quality of surgical training will:</td>
<td>Have an overall positive effect on training.</td>
<td>4.54</td>
<td>0.56</td>
<td>0.00</td>
<td>97.14</td>
</tr>
<tr>
<td></td>
<td>Encourage competition and drive up standards of training.</td>
<td>4.23</td>
<td>0.91</td>
<td>0.00</td>
<td>85.71</td>
</tr>
<tr>
<td></td>
<td>Create accountability for individual trainers.</td>
<td>4.23</td>
<td>0.65</td>
<td>0.00</td>
<td>88.57</td>
</tr>
<tr>
<td></td>
<td>Assist in assigning trainees to posts appropriate to their training needs.</td>
<td>4.20</td>
<td>0.58</td>
<td>0.00</td>
<td>91.43</td>
</tr>
<tr>
<td>How should the results of measuring surgical training quality be used?</td>
<td>Underperforming trusts should be contacted.</td>
<td>4.09</td>
<td>0.82</td>
<td>0.00</td>
<td>85.71</td>
</tr>
<tr>
<td></td>
<td>Should be made available to the Deanery/Training Board.</td>
<td>4.11</td>
<td>0.87</td>
<td>0.00</td>
<td>80.00</td>
</tr>
<tr>
<td>Exposure to operative cases</td>
<td>Case mix is an important marker of surgical training quality.</td>
<td>4.03</td>
<td>0.71</td>
<td>0.00</td>
<td>82.86</td>
</tr>
<tr>
<td></td>
<td>Volume of cases is an important marker of surgical training quality.</td>
<td>3.83</td>
<td>0.75</td>
<td>0.00</td>
<td>80.00</td>
</tr>
<tr>
<td>Operating Room</td>
<td>Operating lists should be structured to take into account time for trainees to operate.</td>
<td>4.26</td>
<td>0.66</td>
<td>0.00</td>
<td>94.29</td>
</tr>
<tr>
<td></td>
<td>Trainees should be involved in all types of operative cases even if complicated.</td>
<td>4.46</td>
<td>0.78</td>
<td>0.00</td>
<td>94.29</td>
</tr>
<tr>
<td></td>
<td>Training lists should incorporate repetition of similar procedures to aid learning.</td>
<td>4.14</td>
<td>0.65</td>
<td>0.00</td>
<td>85.71</td>
</tr>
<tr>
<td>Supervision versus Independence</td>
<td>The level of supervision is an important marker of training quality.</td>
<td>4.20</td>
<td>0.63</td>
<td>0.00</td>
<td>88.57</td>
</tr>
<tr>
<td></td>
<td>The balance between supervision and independence is an important marker of training quality.</td>
<td>4.23</td>
<td>0.60</td>
<td>0.00</td>
<td>91.43</td>
</tr>
<tr>
<td></td>
<td>High quality training facilitates an increasing level of independence for the trainee.</td>
<td>4.63</td>
<td>0.49</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>A stepwise approach to training is a marker of high quality</td>
<td></td>
<td>4.11</td>
<td>0.76</td>
<td>0.00</td>
<td>82.86</td>
</tr>
<tr>
<td>A high quality surgical training post (rotation) should challenge the trainee to make progress during the post</td>
<td></td>
<td>4.57</td>
<td>0.70</td>
<td>0.00</td>
<td>94.29</td>
</tr>
<tr>
<td>Relationship between the trainer and trainee</td>
<td>A good rapport is essential.</td>
<td>4.40</td>
<td>0.74</td>
<td>0.00</td>
<td>91.43</td>
</tr>
<tr>
<td></td>
<td>The trainer must be approachable.</td>
<td>4.40</td>
<td>0.55</td>
<td>0.00</td>
<td>97.14</td>
</tr>
<tr>
<td></td>
<td>The trainer must be willing to allow the trainee to perform operative cases.</td>
<td>4.69</td>
<td>0.53</td>
<td>0.00</td>
<td>97.14</td>
</tr>
</tbody>
</table>
An Evaluation of the Quality of Surgical Training

<table>
<thead>
<tr>
<th>Teamwork</th>
<th>Trainee rapport with other trainees and surgeons affects the quality of surgical training.</th>
<th>4.09</th>
<th>0.56</th>
<th>0.00</th>
<th>88.57</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A high quality surgical training post (rotation) starts with clear objectives set jointly by the trainer and trainee</td>
<td>4.32</td>
<td>0.87</td>
<td>0.03</td>
<td>85.29</td>
</tr>
<tr>
<td>Constructive Feedback</td>
<td>Trainers should offer regular, constructive feedback to the trainee.</td>
<td>4.74</td>
<td>0.51</td>
<td>0.00</td>
<td>97.14</td>
</tr>
<tr>
<td></td>
<td>Trainees should be given feedback about their training ability.</td>
<td>4.63</td>
<td>0.65</td>
<td>0.00</td>
<td>91.43</td>
</tr>
<tr>
<td>Strong engagement of the trainer with the training process is a marker of the quality of surgical training</td>
<td></td>
<td>4.31</td>
<td>0.63</td>
<td>0.00</td>
<td>91.43</td>
</tr>
<tr>
<td>Personalised training</td>
<td>It is essential for training placements to be matched to the individual trainee's level.</td>
<td>4.00</td>
<td>0.87</td>
<td>0.00</td>
<td>80.00</td>
</tr>
<tr>
<td>Structure and Organisation</td>
<td>Departmental structure and organisation can determine surgical training quality.</td>
<td>4.17</td>
<td>0.86</td>
<td>0.00</td>
<td>82.86</td>
</tr>
<tr>
<td>Simulation and Teaching</td>
<td>Presence of a structured teaching programme is a marker of surgical training quality.</td>
<td>4.37</td>
<td>0.69</td>
<td>0.00</td>
<td>88.57</td>
</tr>
<tr>
<td>Trainee feedback</td>
<td>Aggregate measures from a number of trainees should be used.</td>
<td>4.06</td>
<td>0.73</td>
<td>0.00</td>
<td>82.86</td>
</tr>
<tr>
<td></td>
<td>Trainees should provide formative evaluations of their trainers.</td>
<td>4.21</td>
<td>0.68</td>
<td>0.03</td>
<td>91.18</td>
</tr>
<tr>
<td>Multi-source feedback to measure surgical training quality?</td>
<td>Trainee feedback is an essential component.</td>
<td>4.26</td>
<td>0.61</td>
<td>0.00</td>
<td>91.43</td>
</tr>
<tr>
<td></td>
<td>Trainer feedback is an essential component.</td>
<td>4.23</td>
<td>0.55</td>
<td>0.00</td>
<td>94.29</td>
</tr>
<tr>
<td>Trainee improvement</td>
<td>There should be evidence of trainee improvement during each training post.</td>
<td>4.09</td>
<td>0.61</td>
<td>0.00</td>
<td>91.43</td>
</tr>
<tr>
<td></td>
<td>This includes formative evaluations of trainee from trainers</td>
<td>4.26</td>
<td>0.44</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>Formative assessment should include practical technical assessment of trainees.</td>
<td>4.40</td>
<td>0.55</td>
<td>0.00</td>
<td>97.14</td>
</tr>
<tr>
<td></td>
<td>Formative assessment should include assessment of non-technical skills of trainees.</td>
<td>4.51</td>
<td>0.56</td>
<td>0.00</td>
<td>97.14</td>
</tr>
</tbody>
</table>
### Figure 5b: Surgical training Quality Assessment Tool (S-QAT)

<table>
<thead>
<tr>
<th>Trainer – trainee relationship</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapport between trainer and trainee</td>
<td>Poor rapport</td>
<td>Initially poor, improves over time</td>
<td>Excellent rapport</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approachability of the Trainer</td>
<td>Unapproachable, dismissive or intimidating</td>
<td>Largely approachable but can be intimidating at times</td>
<td>Approachable trainer, invites trainee to discussion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trainer attitude towards trainees performing operative cases</td>
<td>Trainer reluctant to allow trainee to perform operative cases even if competent to do so</td>
<td>Trainer permits trainee to perform cases they are competent in, however does not encourage progression of operative experience</td>
<td>Trainer encourages trainee to perform cases they are competent in and to progress onto more complex cases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure to operative cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case Mix</td>
<td>Minimal variety of cases. Cases not appropriate to the learning needs of the trainee</td>
<td>Diverse variety of cases; may not be suited to the learning needs of the trainee</td>
<td>Diverse variety of cases appropriate to the learning needs of the trainee</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume of Cases</td>
<td>Low volume of cases to provide meaningful experience or too many cases to permit training</td>
<td>Sufficient volume of cases but may be too high to permit teaching time for cases</td>
<td>Good balance of volume while allowing enough time to train during the cases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision of operative procedures</td>
<td>Inadequate supervision of operative procedures; trainee often feels unsupported in the operating room (OR)</td>
<td>Adequate supervision of operative procedures; sometimes trainee feels unsupported in the OR</td>
<td>Appropriate supervision of operative procedures, trainee never reports feeling unsupported in the OR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision of perioperative care</td>
<td>Inadequate supervision of perioperative care; trainee often feels unsupported on the wards</td>
<td>Adequate supervision of perioperative care; sometimes trainee feels unsupported on the wards</td>
<td>Appropriate supervision of perioperative care, trainee never reports feeling unsupported on the wards</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision of outpatient clinics</td>
<td>Inadequate supervision in outpatient clinic; trainee often feels unsupported in the clinic</td>
<td>Adequate supervision in outpatient clinic; sometimes trainee feels unsupported in the clinic</td>
<td>Appropriate supervision in outpatient clinic, trainee never reports feeling unsupported in the clinic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision of endoscopy</td>
<td>Inadequate supervision of endoscopy; trainee often feels unsupported in the endoscopy suite</td>
<td>Adequate supervision of endoscopy; sometimes trainee feels unsupported in the endoscopy suite</td>
<td>Appropriate supervision of endoscopic procedures, trainee never reports feeling unsupported in the endoscopy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Independence

<table>
<thead>
<tr>
<th>Balance between supervision and independence in the operating room</th>
<th>Poor balance, either no opportunities for independence in the OR or under supervised</th>
<th>Adequate balance, limited opportunities for independence in the OR, occasional difficulty obtaining supervision when required</th>
<th>Excellent balance, trainee has opportunities for independence in the OR with readily available supervision upon request</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance between supervision and independence during perioperative care</td>
<td>Poor balance, either no opportunities for independent decision making on the wards or under supervised</td>
<td>Adequate balance, limited opportunities for independent decision making on the wards, occasional difficulty obtaining supervision when required</td>
<td>Excellent balance, trainee has opportunities for independent decision making on the wards with readily available supervision upon request</td>
</tr>
<tr>
<td>Balance between supervision and independence in outpatient clinics</td>
<td>Poor balance, either no opportunities for independent decision making in clinics or under supervised</td>
<td>Adequate balance, limited opportunities for independent decision making in clinics, occasional difficulty obtaining supervision when required</td>
<td>Excellent balance, trainee has opportunities for independent decision making in clinics with readily available supervision upon request</td>
</tr>
<tr>
<td>Balance between supervision and independence in endoscopy</td>
<td>Poor balance, either no opportunities for independence during endoscopy or under supervised</td>
<td>Adequate balance, limited opportunities for independence during endoscopy, occasional difficulty obtaining supervision when required</td>
<td>Excellent balance, trainee has opportunities for independence during endoscopy with readily available supervision upon request</td>
</tr>
<tr>
<td>Level of independence for the trainee.</td>
<td>Limited experience of independent working for the trainee</td>
<td>Some experience of working independently but limited progression</td>
<td>Progressive increase in independence of the trainee is facilitated</td>
</tr>
</tbody>
</table>

### Constructive Feedback

<p>| Trainer to trainee feedback – frequency | Feedback not offered | Brief and/or infrequent feedback | Regular feedback at predefined intervals |
| Trainer to trainee feedback – quality | Critical or judgemental feedback without any recommendations for improvement | Constructive feedback, can be judgemental at times | Constructive feedback with clear recommendations for improvement |
| Trainee to trainer feedback about their training ability – frequency | Feedback not offered | Brief and/or infrequent feedback | Regular feedback at predefined intervals |
| Trainee to trainer feedback about their training ability – quality | Critical or judgemental feedback without any recommendations for improvement | Constructive feedback, can be judgemental at times | Constructive feedback with clear recommendations for improvement |</p>
<table>
<thead>
<tr>
<th>Trainer Engagement</th>
<th>Little knowledge or compliance with the curriculum</th>
<th>Adheres to the training curriculum, sometimes unaware of latest developments to the curriculum</th>
<th>Strongly adherent to the curriculum. Participates in the improvement of the curriculum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainer engagement with workplace based assessments</td>
<td>Poorly engaged with workplace based assessments, does not complete assessments</td>
<td>Engages with workplace based assessments, but does not complete assessments in a timely manner</td>
<td>Strongly engaged with workplace based assessments, encourages their use and timely completion</td>
</tr>
<tr>
<td>Structure and Organisation</td>
<td>Poor team structure, members of the team unsure of their roles and responsibilities</td>
<td>Reasonable team structure, sometimes team members are unsure of their role</td>
<td>Excellent team structure, all members of the team have a well defined role</td>
</tr>
<tr>
<td>Departmental organisation</td>
<td>Poorly organised surgery department. Theatre and clinics often changed without prior communication.</td>
<td>Organised surgery department. Theatres and clinics rarely changed without prior communication</td>
<td>Excellent organisation of department with advance notice of all theatre lists and clinics</td>
</tr>
<tr>
<td>Trainee Progress</td>
<td>Minimal opportunity for trainee to demonstrate progress in operative ability and may demonstrate regression in some cases</td>
<td>Challenging for the trainee but only limited progress in operative ability demonstrated</td>
<td>Challenging, allowing the trainee to demonstrate clear evidence of progress in operative ability</td>
</tr>
<tr>
<td>Operative surgery</td>
<td>Minimal opportunity for trainee to demonstrate progress in peri-operative management and may demonstrate regression in some cases</td>
<td>Challenging for the trainee but only limited progress in peri-operative management demonstrated</td>
<td>Challenging, allowing the trainee to demonstrate clear evidence of progress in peri-operative management</td>
</tr>
<tr>
<td>Peri-operative care</td>
<td>Minimal opportunity for trainee to demonstrate progress in out-patient skills and may demonstrate regression in some cases</td>
<td>Challenging for the trainee but only limited progress demonstrated in out-patient skills</td>
<td>Challenging, allowing the trainee to demonstrate clear evidence of progress in out-patient skills</td>
</tr>
<tr>
<td>Out-patient clinic</td>
<td>Minimal opportunity for trainee to demonstrate progress in endoscopic ability and may demonstrate regression in some cases</td>
<td>Challenging for the trainee but only limited progress demonstrated in endoscopic ability</td>
<td>Challenging, allowing the trainee to demonstrate clear evidence of progress in endoscopic ability</td>
</tr>
<tr>
<td>Endoscopy</td>
<td>Minimal opportunity for trainee to demonstrate progress in endoscopic ability and may demonstrate regression in some cases</td>
<td>Challenging for the trainee but only limited progress demonstrated in endoscopic ability</td>
<td>Challenging, allowing the trainee to demonstrate clear evidence of progress in endoscopic ability</td>
</tr>
<tr>
<td>Multi-disciplinary decision making</td>
<td>Minimal opportunity for trainee to demonstrate progress and may demonstrate regression in some cases</td>
<td>Challenging for the trainee but only limited progress demonstrated</td>
<td>Challenging, allowing the trainee to demonstrate clear evidence of progress</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Personalised training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operative surgery</td>
<td>Training placement is not suited to the trainees’ goals, objectives, and operative ability</td>
<td>Training placement is appropriately suited to the trainees’ goals, objectives and operative ability; albeit some cases too advanced or not challenging enough for trainees’ level</td>
<td>Training placement is well suited to the trainees’ goals, objectives and operative ability</td>
</tr>
<tr>
<td>Peri-operative care</td>
<td>Training placement is not suited to the trainees’ goals, objectives, and peri-operative ability</td>
<td>Training placement is reasonably suited to the trainees’ goals, objectives and peri-operative ability; some cases too advanced or not challenging enough for trainees’ level</td>
<td>Training placement is well suited to the trainees’ goals, objectives and peri-operative ability</td>
</tr>
<tr>
<td>Out-patient clinic</td>
<td>Training placement is not suited to the trainees’ goals, objectives, and out-patient skills</td>
<td>Training placement is reasonably suited to the trainees’ goals, objectives and out-patient skills; some cases too advanced or not challenging enough for trainees’ level</td>
<td>Training placement is well suited to the trainees’ goals, objectives and out-patient skills</td>
</tr>
<tr>
<td>Endoscopy</td>
<td>Training placement is not suited to the trainees’ goals, objectives, and endoscopic ability</td>
<td>Training placement is reasonably suited to the trainees’ goals, objectives and endoscopic ability; some cases too advanced or not challenging enough for trainees’ level</td>
<td>Training placement is well suited to the trainees’ endoscopic goals, objectives and endoscopic ability</td>
</tr>
<tr>
<td>Simulation and Teaching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approach to training</td>
<td>Absence of a structured training program</td>
<td>Structured training program present, no opportunity to tailor to trainee’s needs as they progress</td>
<td>Structured and stepwise approach with opportunity to tailor teaching experiences to trainee’s needs</td>
</tr>
<tr>
<td>Knowledge based teaching</td>
<td>Disorganised or absent teaching</td>
<td>Structured teaching</td>
<td>Excellent structure of teaching</td>
</tr>
<tr>
<td>Programme</td>
<td>Programme</td>
<td>Programme, can be disorganised or frequent cancellations</td>
<td>Programme</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Technical skills teaching programme</td>
<td>Absence of a teaching programme and simulation facilities</td>
<td>Structured teaching programme with a simulation facility, can be difficult to access and frequent cancellations</td>
<td>Excellent structure of teaching programme, good use of simulation facilities</td>
</tr>
<tr>
<td>Non-technical skills teaching programme</td>
<td>Disorganised or absent teaching programme, poor use of or absent simulation facilities</td>
<td>Structured teaching programme, can be disorganised or frequent cancellations</td>
<td>Excellent structure of teaching programme, good use of simulation facilities</td>
</tr>
<tr>
<td>Research skills teaching programme</td>
<td>Disorganised or absent teaching programme</td>
<td>Structured teaching programme, can be disorganised or frequent cancellations</td>
<td>Excellent structure of teaching programme</td>
</tr>
</tbody>
</table>

**Teamwork**

<table>
<thead>
<tr>
<th>Trainee rapport with other trainees</th>
<th>Poor rapport, competitive and individualistic environment</th>
<th>Initially poor, improves over time</th>
<th>Excellent rapport, helpful, cooperative environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainee rapport with other surgeons</td>
<td>Poor rapport, trainees confined to working within own teams</td>
<td>Acceptable rapport with surgeons from other teams</td>
<td>Excellent rapport with surgeons from other teams, good cooperation between teams</td>
</tr>
</tbody>
</table>

**Setting learning objectives**

<table>
<thead>
<tr>
<th>Operative learning objectives</th>
<th>No operative objectives set at start of training post e.g. index procedures</th>
<th>Operative objectives set at start of training post, can be vague and non-specific</th>
<th>Clear and specific operative objectives set jointly by trainer and trainee at start of training post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peri-operative learning objectives</td>
<td>No peri-operative objectives set at start of training post</td>
<td>Peri-operative objectives set at start of training post, can be vague and non-specific</td>
<td>Clear and specific peri-operative objectives set jointly by trainer and trainee at start of training post</td>
</tr>
<tr>
<td>Out-patient learning objectives</td>
<td>No out-patient objectives set at start of training post e.g. management algorithms for common conditions</td>
<td>Out-patient objectives set at start of training post, can be vague and non-specific</td>
<td>Clear and specific out-patient objectives set jointly by trainer and trainee at start of training post</td>
</tr>
<tr>
<td>Endoscopic learning objectives</td>
<td>No endoscopic objectives set at start of training post</td>
<td>Endoscopic objectives set at start of training post, can be vague and non-specific</td>
<td>Clear and specific endoscopic objectives set jointly by trainer and trainee at start of training post</td>
</tr>
</tbody>
</table>
### Operating Room

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Schedule</td>
<td>No provisions for extra time for trainees to operate</td>
</tr>
<tr>
<td>Time allowed for trainees to operate</td>
<td>Structured schedule to allow time for trainees to operate</td>
</tr>
<tr>
<td>Trainee involvement in operative cases</td>
<td>Limited to simple or small proportion of cases regardless of trainee experience</td>
</tr>
<tr>
<td>Involvement in most cases</td>
<td>Involved in all cases including supervised components of complex cases</td>
</tr>
<tr>
<td>Presence of designated operating lists designed for training rather than service</td>
<td>Few or absent, limited or no repetition of similar procedures</td>
</tr>
<tr>
<td>Regular training lists, infrequent repetition of similar procedures</td>
<td>Regular and frequent training lists, repetition of similar procedures to aid learning</td>
</tr>
</tbody>
</table>

### Training Evaluation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainee Feedback</td>
<td>Individual feedback is used to inform decisions and identity of trainee is known</td>
</tr>
<tr>
<td>Aggregate feedback is used to inform decisions, trainee’s identity may be inferred</td>
<td>Anonymised and aggregated feedback, trainee’s identity cannot be inferred</td>
</tr>
<tr>
<td>Trainee evaluation of trainers</td>
<td>No opportunity for trainee feedback to trainer</td>
</tr>
<tr>
<td>Evaluations performed, however unstructured approach</td>
<td>Structured evaluations are performed for majority of trainers encountered</td>
</tr>
<tr>
<td>Multi-source feedback</td>
<td>Neither trainee feedback nor trainee feedback is collated</td>
</tr>
<tr>
<td>Either trainee feedback or trainer feedback is collated</td>
<td>Both trainee and trainer feedback is collated</td>
</tr>
</tbody>
</table>

### Trainee Assessment

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence of trainee improvement during a training post</td>
<td>Little or no evidence, or evidence of deterioration</td>
</tr>
<tr>
<td>Limited evidence, or evidence shows minimal improvement</td>
<td>Robust evidence of notable improvement</td>
</tr>
<tr>
<td>Formative evaluation of trainee from trainers</td>
<td>Little or no assessment</td>
</tr>
<tr>
<td>Structured assessment, limited evidence of improvement</td>
<td>Structured and regular assessment, clear evidence of improvement</td>
</tr>
<tr>
<td>Technical skills assessment</td>
<td>Little or no assessment</td>
</tr>
<tr>
<td>Structured assessment, limited evidence of improvement</td>
<td>Structured and regular assessment, clear evidence of improvement</td>
</tr>
<tr>
<td>Non-technical skills assessment</td>
<td>Little or no assessment</td>
</tr>
<tr>
<td>Structured assessment, limited evidence of improvement</td>
<td>Structured and regular assessment, clear evidence of improvement</td>
</tr>
</tbody>
</table>

### Quality Control

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collation of training quality information</td>
<td>No collection of information about training quality</td>
</tr>
<tr>
<td>Some collection of information about training quality, no</td>
<td>Organised collection of training quality data, made available to</td>
</tr>
<tr>
<td>Action on training quality information</td>
<td>formal structure in place</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Little or no action</td>
<td>Underperforming hospitals and departments contacted</td>
</tr>
</tbody>
</table>
There was consensus that measuring the quality of surgical training will have a positive effect on training. Most experts believed this would encourage competition and drive up the standards of training. Complete transparency of training quality outcomes failed to reach consensus. Instead, consensus was reached that the results of measuring training quality should be used to contact underperforming training units and that the results should be made available to training boards.

Having established consensus that there is a need and a purpose for measuring the quality of surgical training, this chapter went on to define high quality surgical training. As would be expected, exposure to operative cases was identified as a marker of training quality. However, whilst volume of cases reached consensus, case mix was also identified as an important marker of surgical training quality. For example, a good mix of minimally invasive or laparoscopic surgery alongside open surgery. There was consensus that while supervision is an important indicator of quality, this needs to be balanced by a progressive increase towards independence. High quality training should facilitate an increasing level of independence for the trainee and should challenge the trainee to make progress during the post.

Some important non-technical aspects of training were identified as markers of quality. These included the relationships between the trainee and the trainer but also the relationships with other members of the team. For example, senior trainees can potentially mentor junior trainees. A high quality surgical training post should start with clear objectives set jointly by the trainer and trainee. Trainers should offer regular, constructive feedback to the trainee. This can
be in a combination of informal or written feedback. Importantly consensus was reached that trainers should also receive feedback about their training ability.

Efficient matching of training placements to the individual trainees may maximize the training opportunities available. For instance, a junior trainee is unlikely to be suited to an advanced minimally invasive surgical subspecialty placement and this would be more suited to a senior resident or even a sub-specialty fellow. Structural aspects of a training position such as shift patterns can affect the continuity between the trainee and trainer and subsequently affect the rapport. Presence of a structured teaching programme reached consensus as a marker of surgical training quality. However, the role of simulation within this teaching programme was less well defined. It appears that the structure of the teaching programme and its delivery is a better marker of quality than simply the possession of equipment.

When measuring training quality, aggregate measures were favoured. The consensus was that multi-source feedback should be used to measure surgical training quality. Trainer feedback is an essential component as they may know the limitations of their training program and may appreciate the opportunity to try and improve them. Training is a multi-faceted process and its ultimate success should be judged by actual evidence of trainee improvement during each training post. The consensus was that formative assessments should be performed. Interestingly written assessments did not reach consensus. Instead practical technical and non-technical assessments of trainees were favoured.
5.4: Discussion

Using modified Delphi methodology, an international panel of experts in surgical education have reached consensus on how to define and measure the processes for high quality surgical training. Items reaching consensus have been used to develop the Surgical training Quality Assessment Tool (S-QAT). Descriptive anchors have been developed to assign levels to each point on the quality scale. This tool forms the first step towards robust evaluation of surgical training quality, in real time. The results of this chapter support robust evaluation of training quality in order to raise overall standards of surgical training quality.

Through a combination of trainers’ and trainees’ opinions and the consensus process using the international panel of surgical education experts, the assessment of training quality naturally adopted Donabedian’s 3-dimensional conceptual framework: structure, process and outcomes (Donabedian 1988). Surgical training quality can be assessed by examining the organisation in which it is delivered, thus examining the structure. Examples of structure include the volume and case mix of surgical procedures performed at an institution, the presence of a structured training programme and dedicated training operating lists. The training techniques, including the trainers themselves and the quality of supervision and constructive feedback they provide, assess the process. The ultimate outcomes are the improvement in performance of the trainees. Measuring trainee performance is in itself a challenging area. The ideal barometers of trainee performance would be the patient healthcare outcomes and patient experiences of the patients treated by the trainees. As with other forms of surgical quality measurement, continuous evolvement of the S-QAT tool will likely be
necessary. This first step looks largely at the process and the challenge in the future will be to define objective endpoints that can be used as outcomes for program evaluation. Furthermore, as is the case for most strategies to measure quality in surgery, it will likely involve using composite indices as individual metrics used on their own may only have a limited correlation with the overall surgical training quality (Merkow et al. 2013).

Given that most experts believed measuring training quality would encourage competition and drive up the standards of training, it was surprising that complete transparency of training quality outcomes failed to reach consensus. This likely reflects a cautious attitude towards summative assessment of training quality that may lead to high stakes decisions such as accreditation or in terms of attracting the best residents to a program. In addition, aggregate measures were favoured. This avoids the undue influence of a single good or bad experience and helps to anonymise feedback from individuals, therefore resulting in a more honest assessment. There was an appetite for formative assessment such as identifying areas for improvement in underperforming units and providing them with information on how to optimise these areas. In support of this there was consensus that trainers should receive feedback about their training ability. This would provide them with the opportunity to refine and improve their training techniques. This is consistent with the growing literature on the benefits of training the trainer courses and interventions (Hull et al. 2013; Murphy et al. 2008).

With the introduction of a novel quality scale, the focus often lies in identifying poor performers. Indeed the positive effects of competition may be associated with punitive
measures given to the ‘low performers’. However, punitive measures are not necessarily the most effective way in which competition and feedback can drive up standards. Areas of best practice should be identified and the information used for continuous improvement to training programs through a process of rigorous formative feedback. Although the measurement of quality in surgical training is novel, assessing quality of higher education institutes is well established (USN 2012). Within surgery, the evaluation of quality in clinical outcomes has been strong a driver for raising clinical standards. The introduction of the National Surgical Quality Improvement Program was hailed as the most comprehensive comparative clinical outcomes assessment program with the objective of enhancing the quality of surgical care (Khuri et al. 1998). There is evidence that the use of NSQIP can improve the quality of clinical care (Compoginis and Katz 2013; Hall et al. 2009). It must be acknowledged that NSQIP uses risk-adjusted outcomes, while the S-QAT tool focuses on process measures and objective outcomes require further development. Introduction of objective evaluation of the quality of surgical training programs could similarly act as a driver to improve surgical training outcomes globally. The UK’s Joint Committee on Specialist Training (JCST) report quality indicators for training placements (JCST 2012a). However the development of these quality indicators is not explained and the relevance of some indicators is debatable e.g. trainees ‘should participate in operative briefings...’. Indicators such as those from the JCST can also be inferred from the Accreditation Council for Graduate Medical Education Surgical Residency program requirements (ACGME 2012). While the JCST trainee survey is an encouraging step towards the assessment of training quality, it only samples the subjective views of trainees. This is a significant limitation as there is evidence that when asked to define service and education, trainees’ and trainers’ perspectives
demonstrated significant differences (Sanfey et al. 2011). In this chapter a systematic and evidence-based approach was employed to develop a tool to enable evaluation of surgical training quality. The consensus was that multi-source feedback including both the trainees (consumers of training) and the consultant trainers (providers of training) is required.

The chapter has some limitations that should be acknowledged. Patients’ opinions were not sampled in this study. However, patients’ lack of detailed knowledge regarding the delivery of training would make them unable to make an informed contribution to the study. The semi-structured interviews conducted to develop themes for the Delphi study in chapter 4, were all conducted in a single country. This pragmatic approach was employed as it was felt that face-to-face semi-structured interviews would yield more information than online or audio interviews. In addition, as the focus was on ideal training programs and not analysis of the current training programs it was felt that the themes developed would be applicable to an international community. A potential limitation of using an international panel of experts for the Delphi process is that the tool may not be sensitive to local concerns that can vary in different countries. The tool may benefit from adaptation for use in different countries, particularly if involving a different language, in a similar way that other assessment tools have been culturally adapted for use in different healthcare systems (Passauer-Baierl et al. 2014). There were some missing data points where members of the Delphi expert panel did not rate all the statements. However, overall there were only 7 missing data points in the final round (0.286%), and Cronbach’s alpha remained over 0.930 whether missing data points were replaced with means, modes or 3 (neutral). Future studies will endeavour to pilot the use of the
tool to allow further refinement and assess its usability and reliability. The next step will be to develop objective outcome measures to complement the structure and process measures assessed by the newly developed S-QAT. The final step will be to develop appropriately weighted composite measures to assess surgical training quality.

Educational commissioning is undergoing change in the UK where there is agreement that it may not be suitable for all consultants and hospitals to be assigned surgical trainees (DoH 2012). Similarly in the USA the 2011 ACGME Duty Hour Standards go beyond limits on work hours in order to promote high quality education, with particular attention paid to graduated standards for resident supervision (ACGME 2011b). Hospitals and surgical training programs will need to demonstrate high quality training in order to receive trainees and maintain accreditation. The S-QAT tool developed in this chapter could enable hospitals and surgical training programs to demonstrate high quality surgical training and to monitor standards longitudinally. Furthermore the formative feedback provided will enable programs to identify their strengths and weaknesses and subsequently initiate quality improvement measures.

5.5: Conclusion

There is a need to robustly assess surgical training quality. International expert consensus was achieved on the definition of high quality surgical training and how it can be measured. This has been developed into the Surgical training Quality Assessment Tool (S-QAT) that can be used to evaluate surgical training programs globally. There was almost unanimous agreement that
competition created by measuring training quality will raise standards of surgical education. Future studies will need to pilot the use of the tool to allow further refinement and assess its usability and reliability.
Chapter 6: Pilot Implementation of a Surgical Training

Quality Assessment Tool
6.1: Peer Review of the Surgical Training Quality Assessment Tool

6.1.1: Introduction

Existing evidence suggests that the quality of training received can determine patient outcomes (Asch et al. 2009). Variations in global general surgical training curricula have been evaluated in chapter 2. Chapter 3 confirmed significant variations in surgical training programs when viewed from the perspective of a surgical trainee. While efforts to standardize the surgical curriculum may lead to better quality of training, our explicit knowledge of what constitutes a high quality surgical training program is limited. Further, on-going pressures to reduce training hours, necessitate that we identify innovative, pragmatic, and evidence-based methods to maximise the quality of training and define the best assessment methods of our trainees.

In chapter 4 an interview study sampling both trainees and trainers identified indicators of high quality surgical training. Using the information gathered, in chapter 5 a Delphi consensus method was employed to develop international expert consensus on the indicators of high quality surgical training. The results of chapter 5 were used to develop the S-QAT Surgical Training Quality Assessment Tool. The objective of this chapter is to further develop the S-QAT by initially subjecting it to an open peer review process before pilot testing the tool.
Figure 6a: Written Feedback Questionnaire

APDS Workshop – Defining High Quality Surgical Training

1. In your opinion, currently what are the three most important issues with the quality of surgical training?
   i. ________________________________
   ii. ________________________________
   iii. ________________________________

2. Surgical training quality should be measured?
   Strongly Disagree 1 2 3 4 5 Strongly Agree

3. The S-QAT: Surgical training Quality Assessment Tool, will capture variation in surgical training quality offered by residency programs?
   Strongly Disagree 1 2 3 4 5 Strongly Agree

4. How should the S-QAT be used? (you may agree or disagree with multiple statements)
   As a questionnaire to residents
     Strongly Disagree 1 2 3 4 5 Strongly Agree
   As a questionnaire to training faculty
     Strongly Disagree 1 2 3 4 5 Strongly Agree
   By independent assessors of a residency program
     Strongly Disagree 1 2 3 4 5 Strongly Agree
   As formative feedback system for residency programs
     Strongly Disagree 1 2 3 4 5 Strongly Agree
   As summative feedback system for residency programs
     Strongly Disagree 1 2 3 4 5 Strongly Agree
   Other ________________________________

5. Would you consider using the S-QAT?
   Strongly Disagree 1 2 3 4 5 Strongly Agree

6. How could the S-QAT be improved?
   i. ________________________________
   ii. ________________________________
   iii. ________________________________

Please continue overleaf if necessary

7. Would you like to be involved in further efforts to develop & pilot the S-QAT? (personal details are optional if you reply no).
   a. No
   b. Yes
     i. Name (optional) ________________________________
     ii. Email address (optional) ________________________________
     iii. Institution (optional) ________________________________
     iv. Position (optional) ________________________________

On behalf of all the presenters I would like to thank you for attending and contributing to this Workshop.

Please do not hesitate to contact me should you have any further questions or recommendations.

Pritam

Dr Pritam Singh MA MBBS MRCs
Clinical Research Fellow
Imperial College London
pritam.singh@imperial.ac.uk
6.1.2: Methods

An interactive workshop was proposed for the Association of Program Directors in Surgery, Surgical Education Week in April 2013, Orlando, Florida, USA. The format for the interactive workshop was to initially introduce the S-QAT Surgical Training Quality Assessment Tool and follow this by a series of related talks on the subject of measuring training quality by an international team of surgeon educators. The second half of the workshop was run as an interactive discussion with the audience. In addition to verbal feedback during the discussion, attendees were invited to complete written feedback questionnaires (Figure 6a).

6.1.3: Results

The interactive workshop was attended by a total of twenty subjects. Completed written feedback forms were received from ten attendees. The results are displayed in Table 6a. The majority of subjects strongly agreed that surgical training quality should be measured. A small majority agreed that the S-QAT would capture the variation in surgical training quality offered by surgical training programs. One subject suggested that this could be facilitated if the S-QAT was combined with other measurement tools. When asked how the S-QAT should be used, there was some variation in responses with a small majority agreeing that it should be used as a questionnaire to trainees and trainers. A large majority felt the S-QAT should be used as formative feedback for surgical training programs. There were mixed feelings for S-QAT’s use as summative feedback. A small majority of subjects indicated they would consider using the S-QAT.
### Table 6a: Written Feedback from the Interactive Workshop

<table>
<thead>
<tr>
<th>Question</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 - Strongly Disagree</td>
</tr>
<tr>
<td>Surgical training quality should be measured?</td>
<td>1</td>
</tr>
<tr>
<td>The S-QAT: Surgical training Quality Assessment Tool, will capture variation in surgical training quality offered by residency programs?</td>
<td>2</td>
</tr>
<tr>
<td>How should the S-QAT be used?</td>
<td>1</td>
</tr>
<tr>
<td>* As a questionnaire to residents</td>
<td>1</td>
</tr>
<tr>
<td>‘Too many already’</td>
<td>1</td>
</tr>
<tr>
<td>* As a questionnaire to training faculty</td>
<td>1</td>
</tr>
<tr>
<td>‘Too many already’</td>
<td>1</td>
</tr>
<tr>
<td>* By independent assessors of a residency program</td>
<td>1</td>
</tr>
<tr>
<td>* As formative feedback system for residency programs</td>
<td>1</td>
</tr>
<tr>
<td>* As summative feedback system for residency programs</td>
<td>1</td>
</tr>
<tr>
<td>Would you consider using the S-QAT?</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 6b: Written Feedback from the Interactive Workshop – Free text responses

<table>
<thead>
<tr>
<th>Question</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the three most important issues with the quality of surgical training?</td>
<td>• Balance between autonomy and supervision</td>
<td>• Operative case mix</td>
<td>• Instilling clinical judgment</td>
</tr>
<tr>
<td></td>
<td>• Tools to measure</td>
<td>• Honest feedback</td>
<td>• Difference between residents</td>
</tr>
<tr>
<td></td>
<td>• Consistent knowledge and performance post training</td>
<td>• Known limitations of surgeons post residency</td>
<td>• Volume of types of procedures for competency and maintenance of competency</td>
</tr>
<tr>
<td></td>
<td>• Unusable assessment tool</td>
<td>• Inter-rater variability</td>
<td>• Definition of minimal competency</td>
</tr>
<tr>
<td></td>
<td>• Cynical culture in many programs (e.g. back-stabbing, bad mouthing other professionals)</td>
<td>• Lack of shared mental models among faculty related to training priorities</td>
<td>• Assessment of patient outcomes, safety etc. not integrated into training program and assessment</td>
</tr>
<tr>
<td></td>
<td>• Various metrics pulling them from patients</td>
<td>• Over regulation</td>
<td>• Focus on economics</td>
</tr>
<tr>
<td></td>
<td>• Faculty Development</td>
<td>• No formal feedback to resident or attending</td>
<td>• General discussion of indications and options</td>
</tr>
<tr>
<td></td>
<td>• Unstructured</td>
<td>• Ability to improve if quality deemed substandard</td>
<td>• Regulations imparted to address training quality</td>
</tr>
<tr>
<td></td>
<td>• Support for, obtaining resources</td>
<td>• Lack of agreed quality metrics</td>
<td>• Disparity between goals for training and expected skills at completion</td>
</tr>
<tr>
<td></td>
<td>• Lack of standardization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How could the S-QAT be improved?</td>
<td>• Simpler form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Need more specific definitions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Add situational dimensions to questions (e.g. trainer-trainee relations – in OR, on wards, in conferences)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Some quality indicators need clear definitions (some indicators are vague)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Shorten (takes too long)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• ?consider frames of reference – more detailed to minimize subjectivity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Competency based – (more clear)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Incorporate objective patient outcome data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Too long (too many items)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Simplicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Method to determine reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Is teaching consistent with clinical test evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Better operational definition/anchoring of scale points</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Decrease redundancy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The free-text responses are provided in Table 6b. The lack of tools available to assess surgical training quality was highlighted as an important issue by a number of subjects. Feedback on the S-QAT focused on its usability with a number of subjects commenting on the need to make it simpler and shorter. A method suggested to improve the S-QAT was to incorporate objective patient outcomes data.

6.1.4: Discussion

This chapter describes the open peer-review of the S-QAT Surgical training Quality Assessment Tool. The S-QAT was introduced at an international Surgical Education Conference and both open verbal and written feedback was received. The lack of tools available to assess surgical training quality was highlighted as an important issue. The majority of attendees strongly agreed that surgical training quality should be measured. There was support for the use of S-QAT as a questionnaire to both trainees and trainers. A majority of attendees felt that the S-QAT could capture variation in training quality and should be used for formative feedback to training programs. While summative feedback using the S-QAT received some support, there was also some strong opposition to this. Feedback on the S-QAT itself seemed to focus on the theme of usability with advice to shorten and simplify the S-QAT tool where possible.

The feedback received on the S-QAT was encouraging with some clear suggestions for improvement. The feedback confirmed the conclusions of chapters 4 and 5, which demonstrated enthusiasm to measure surgical training quality amongst the surgical profession. It also confirmed the cautious approach to the implementation of this concept, with support for
S-QAT results to be used as formative feedback but only limited support for its use as summative feedback to surgical training programs. This is likely to be due to the need for further refinement of the tool before its potential use in high stakes decisions. While this workshop included delegates from an international community, the limited numbers should be acknowledged. In an open discussion limited numbers can be advantageous however the generalisability of the discussion will also be limited. Furthermore the subjects in the workshop had only a limited time to study the S-QAT tool and did not actually put it to use by completing it.

The logical next step for the refinement of the S-QAT is to pilot its use based on the feedback received from this Workshop. The support for its use centred on using it as a questionnaire for trainers and trainees and using the results to provide feedback in a formative manner to surgical training programs.

6.2: A Regional Quality Improvement Initiative for Surgical Training

6.2.1: Introduction

Chapters 4 and 5 described the development of a tool to assess the surgical training quality provided by surgical training programs. The first section of this chapter subjected this S-QAT to peer review at an international surgical education conference. The remainder of this chapter further assesses the S-QAT by piloting its use in the form of a quality improvement initiative to
identify variation in training quality across Training Centers in one region. The objectives of this quality improvement initiative were:

- To provide a formative assessment of the quality of surgical training in training centers.
- To identify areas of best practice in surgical training centers.
- To identify areas of variation in surgical training centers.
- Make recommendations to improve the overall quality of surgical training in North West London.

Identification of variations in surgical training quality could be the first step towards improving quality across the region. This could be through identification of best practices and subsequent dissemination to other training centers. Centers where weaknesses are identified could benefit from enhanced support to raise standards. Based on the feedback from the first half of this chapter, a decision was made to provide formative feedback rather than any summative feedback to surgical training programs.

**6.2.2: Methods**

This study involved collaboration with the higher surgical training committee in North West London. All General Surgery training centers in the North West London training region were identified (n=13). Within each training center, all consultant faculty trainers (n=96, median 8,
range 1-12 per center) and general surgical registrars (n=51, median 4, range 1 to 8 per center) were identified. The S-QAT questionnaire was developed into an online questionnaire to facilitate data collection and storage. Quality markers were rated on a 5-point Likert scale. The Chair of the higher surgical training committee electronically distributed invitations to the study to all identified consultant faculty trainers and general surgical registrars via email.

Background information to the study was provided in the email invitation. All subjects were informed that the higher surgical training committee would hold any data collected and any dissemination of information would be anonymised so individuals’ responses would not be identifiable. Subjects were encouraged to support the training quality improvement initiative however there was no obligation to participate and no financial or other rewards were offered to encourage responses.

Invitations to consultant faculty trainers and trainee general surgery registrars were staggered to minimise conferring between trainer and trainee during completion of the S-QAT questionnaire. Electronic invitations contained a hyperlink to the online questionnaire. Consultant faculty trainers were sent the invitations first followed by two reminder emails in an attempt to maximise responses. The process was then repeated for the trainee general surgery registrars.

Data was collected and stored in a secure password protected online questionnaire system (http://www.surveymonkey.com). Data was analysed using Microsoft Excel. High and low
quality ratings were calculated using the means of the responses. Variability was analysed by comparing standard deviations of the responses. Where relevant, responses were separated into consultant faculty trainer responses, trainee general surgery registrar responses and overall responses. Statistical comparisons between the responses of faculty trainers and trainee registrars were performed using the non-parametric Mann-Whitney U test.

6.2.3: Results

Twelve (92.3%) Training Centers responded. Overall sixty-nine (46.9%) responses were received (median 3.5, range 1 to 9 from centers that responded). Forty-five (46.9%) Faculty Trainers responded (median 3.5, range 1 to 8 from centers that responded). Twenty-four (47.1%) trainees responded (median 2, range 1 to 5 from centers that responded). Once the S-QAT questionnaire was started, the overall completion rate was 72.5% (71.1% for faculty trainers and 75% for trainees). Faculty experience ranged from 7 months to 27 years as consultant with a median of 10 years. Trainee experience ranged from ST3 (year 1 of higher surgical registrar training) to ST8 (final year of surgical registrar training) with a median of ST5.

Areas of highest quality included; trainee supervision, trainer approachability and attitude towards trainees performing operative cases, trainee rapport with the trainer and other trainees and the balance between supervision and independence in the perioperative and outpatient setting. Table 6c describes the ten highest quality areas as identified by the S-QAT responses. Table 6d describes the ten lowest quality areas as identified by the S-QAT responses. Areas identified as having potential for improvement in training quality included; the
presence of designated training operating lists, research and non-technical skills teaching programmes, training evaluation and quality control measures.

‘Feedback needs to be specific if it is to be useful. GMC survey is too generic to identify problem and correct it. Trainees must be responsible for the feedback they give’

Areas of greatest variation in training quality included; provision of designated training operating list, out-patient learning objectives, structure of all teaching programmes, multiple aspects of endoscopic training and trainee evaluation of trainers. Table 6e describes the ten highest variability areas as identified by the S-QAT responses. Free text analysis reiterated concerns with endoscopy:

‘Getting endoscopy experience is an on-going issue in my training region - not all posts offer it’

‘I am trying to get my trainee an endoscopy slot though this is proving problematic, due to gastro consultants not providing lists’

Table 6f describes the ten lowest variability areas as identified by the S-QAT responses. Areas of lowest variation in training quality included; supervision in the operating room, rapport between and approachability of the trainer to the trainee, the level of independence for the trainee and trainer to trainee feedback.
There was good concordance between faculty trainer and trainee registrar responses with fifty-one of the fifty-five (93%) quality elements demonstrating no significant differences between the two groups. Faculty trainers rated the following elements of training quality significantly higher than trainee registrars; exposure to operative cases in terms of case mix (mean 4.47 versus 3.86, p = 0.032), trainee involvement in operative cases (mean 4.34 versus 3.77, p = 0.035), trainer attitude towards trainees performing operative cases (mean 4.80 versus 4.27, p = 0.017) and the balance between supervision and independence in outpatient clinics (mean 4.74 versus 4.14, p = 0.005).

Free text analysis revealed a common theme that the S-QAT could be improved by reducing its length. There was also a feeling amongst trainers that some of the questions were more suited to trainees. For example in relation to the question of a trainer’s approachability:

‘Can’t answer - I am the trainer - hope I am approachable’

Faculty trainers demonstrated good insight into the training quality their posts offered:

‘For a junior trainee my current case mix is ok - for a more senior one there is not enough’

‘For a very experienced surgeon, the allocation of very junior trainees (albeit of very high quality) limits the potential advantage of training intermediate level
trainees who could progress through more complicated elements of operative surgery.’

‘Access to simulation is poor’

‘We have extensive research training on offer, but it is directed at research fellows, not clinical trainees, most of whom have already completed MD/PhD’
**Table 6c: High Quality Training As Rated by S-QAT responses**

<table>
<thead>
<tr>
<th>_aspect</th>
<th>Overall</th>
<th>Faculty Trainers</th>
<th>Trainee Registrars</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Mean</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Supervision of endoscopy*</td>
<td>4.58</td>
<td>1.11</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Trainer attitude towards trainees performing operative cases</td>
<td>4.58</td>
<td>0.74</td>
<td>4.80</td>
<td>0.48</td>
</tr>
<tr>
<td>Approachability of the trainer</td>
<td>4.57</td>
<td>0.74</td>
<td>4.76</td>
<td>0.51</td>
</tr>
<tr>
<td>Supervision of operative procedures*</td>
<td>4.57</td>
<td>0.65</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Rapport between trainer and trainee</td>
<td>4.56</td>
<td>0.65</td>
<td>4.60</td>
<td>0.60</td>
</tr>
<tr>
<td>Balance between supervision and independence in outpatient clinics</td>
<td>4.52</td>
<td>0.73</td>
<td>4.38</td>
<td>0.74</td>
</tr>
<tr>
<td>Trainee rapport with other trainees</td>
<td>4.51</td>
<td>0.87</td>
<td>4.73</td>
<td>0.44</td>
</tr>
<tr>
<td>Level of independence for the trainee.</td>
<td>4.49</td>
<td>0.71</td>
<td>4.59</td>
<td>0.65</td>
</tr>
<tr>
<td>Balance between supervision and independence during perioperative care</td>
<td>4.47</td>
<td>0.76</td>
<td>4.59</td>
<td>0.73</td>
</tr>
<tr>
<td>Supervision of perioperative care*</td>
<td>4.43</td>
<td>0.77</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

*Faculty trainers were not asked questions relating to supervision as these were felt to be more appropriate for trainees to answer.*
### Table 6d: Potential Areas for Training Quality Improvement As Rated by S-QAT responses

<table>
<thead>
<tr>
<th>Area</th>
<th>Overall Mean</th>
<th>Overall Standard Deviation</th>
<th>Faculty Trainers Mean</th>
<th>Faculty Trainers Standard Deviation</th>
<th>Trainee Registrars Mean</th>
<th>Trainee Registrars Standard Deviation</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of designated operating lists designed for training rather than service</td>
<td>3.09</td>
<td>1.54</td>
<td>3.21</td>
<td>1.63</td>
<td>2.90</td>
<td>1.38</td>
<td>0.471</td>
</tr>
<tr>
<td>Research skills teaching programme</td>
<td>3.24</td>
<td>1.38</td>
<td>3.29</td>
<td>1.31</td>
<td>3.18</td>
<td>1.46</td>
<td>0.849</td>
</tr>
<tr>
<td>Training Evaluation - Trainee evaluation of trainers</td>
<td>3.37</td>
<td>1.26</td>
<td>3.43</td>
<td>1.10</td>
<td>3.28</td>
<td>1.45</td>
<td>0.705</td>
</tr>
<tr>
<td>Training Evaluation - Trainee Feedback</td>
<td>3.45</td>
<td>1.26</td>
<td>3.28</td>
<td>1.25</td>
<td>3.71</td>
<td>1.23</td>
<td>0.351</td>
</tr>
<tr>
<td>Quality Control - Collation of training quality information</td>
<td>3.48</td>
<td>1.18</td>
<td>3.48</td>
<td>1.24</td>
<td>3.47</td>
<td>1.09</td>
<td>0.850</td>
</tr>
<tr>
<td>Non-technical skills teaching programme</td>
<td>3.50</td>
<td>1.39</td>
<td>3.57</td>
<td>1.40</td>
<td>3.39</td>
<td>1.38</td>
<td>0.584</td>
</tr>
<tr>
<td>Quality Control - Action on training quality information</td>
<td>3.51</td>
<td>1.23</td>
<td>3.64</td>
<td>1.19</td>
<td>3.31</td>
<td>1.26</td>
<td>0.439</td>
</tr>
<tr>
<td>Trainee to trainer feedback about their training ability – frequency</td>
<td>3.58</td>
<td>1.15</td>
<td>3.67</td>
<td>1.07</td>
<td>3.45</td>
<td>1.24</td>
<td>0.564</td>
</tr>
<tr>
<td>Training Evaluation - Multi-source feedback</td>
<td>3.59</td>
<td>1.26</td>
<td>3.57</td>
<td>1.28</td>
<td>3.63</td>
<td>1.22</td>
<td>1.00</td>
</tr>
<tr>
<td>Out-patient learning objectives</td>
<td>3.64</td>
<td>1.32</td>
<td>3.85</td>
<td>1.19</td>
<td>3.32</td>
<td>1.43</td>
<td>0.175</td>
</tr>
</tbody>
</table>
### Table 6e: High Variability in Training Quality As Rated by S-QAT responses

<table>
<thead>
<tr>
<th>Section</th>
<th>Overall Mean</th>
<th>Overall Standard Deviation</th>
<th>Faculty Trainers Mean</th>
<th>Faculty Trainers Standard Deviation</th>
<th>Trainee Registrars Mean</th>
<th>Trainee Registrars Standard Deviation</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of designated operating lists designed for training rather than service</td>
<td>3.09</td>
<td>1.54</td>
<td>3.21</td>
<td>1.63</td>
<td>2.90</td>
<td>1.38</td>
<td>0.471</td>
</tr>
<tr>
<td>Endoscopic learning objectives</td>
<td>3.88</td>
<td>1.49</td>
<td>3.79</td>
<td>1.47</td>
<td>4.00</td>
<td>1.52</td>
<td>0.530</td>
</tr>
<tr>
<td>Technical skills teaching programme</td>
<td>3.67</td>
<td>1.48</td>
<td>3.46</td>
<td>1.57</td>
<td>4.00</td>
<td>1.25</td>
<td>0.252</td>
</tr>
<tr>
<td>Non-technical skills teaching programme</td>
<td>3.50</td>
<td>1.39</td>
<td>3.57</td>
<td>1.40</td>
<td>3.39</td>
<td>1.38</td>
<td>0.584</td>
</tr>
<tr>
<td>Research skills teaching programme</td>
<td>3.24</td>
<td>1.38</td>
<td>3.29</td>
<td>1.31</td>
<td>3.18</td>
<td>1.46</td>
<td>0.849</td>
</tr>
<tr>
<td>Out-patient learning objectives</td>
<td>3.64</td>
<td>1.32</td>
<td>3.85</td>
<td>1.19</td>
<td>3.32</td>
<td>1.43</td>
<td>0.175</td>
</tr>
<tr>
<td>Balance between supervision and independence in endoscopy</td>
<td>4.07</td>
<td>1.28</td>
<td>3.94</td>
<td>1.27</td>
<td>4.30</td>
<td>1.27</td>
<td>0.294</td>
</tr>
<tr>
<td>Trainee Progress - Endoscopy</td>
<td>4.04</td>
<td>1.27</td>
<td>3.93</td>
<td>1.33</td>
<td>4.20</td>
<td>1.17</td>
<td>0.611</td>
</tr>
<tr>
<td>Training Evaluation - Trainee evaluation of trainers</td>
<td>3.37</td>
<td>1.26</td>
<td>3.43</td>
<td>1.10</td>
<td>3.28</td>
<td>1.45</td>
<td>0.705</td>
</tr>
<tr>
<td>Knowledge based teaching programme</td>
<td>3.87</td>
<td>1.26</td>
<td>3.81</td>
<td>1.39</td>
<td>3.95</td>
<td>1.05</td>
<td>0.906</td>
</tr>
</tbody>
</table>
### Table 6f: Low Variability in Training Quality As Rated by S-QAT responses

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Faculty Trainers</th>
<th>Trainee Registrars</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Mean</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Supervision of operative procedures*</td>
<td>4.57</td>
<td>0.65</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Rapport between trainer and trainee</td>
<td>4.56</td>
<td>0.65</td>
<td>4.60</td>
<td>0.60</td>
</tr>
<tr>
<td>Trainee Progress - Out-patient clinic</td>
<td>4.37</td>
<td>0.68</td>
<td>4.48</td>
<td>0.62</td>
</tr>
<tr>
<td>Level of independence for the trainee.</td>
<td>4.49</td>
<td>0.71</td>
<td>4.59</td>
<td>0.65</td>
</tr>
<tr>
<td>Balance between supervision and independence in outpatient clinics</td>
<td>4.52</td>
<td>0.73</td>
<td>4.74</td>
<td>0.50</td>
</tr>
<tr>
<td>Approachability of the trainer</td>
<td>4.57</td>
<td>0.74</td>
<td>4.76</td>
<td>0.51</td>
</tr>
<tr>
<td>Trainer to trainee feedback – frequency</td>
<td>4.25</td>
<td>0.74</td>
<td>4.40</td>
<td>0.66</td>
</tr>
<tr>
<td>Trainee Progress - Operative surgery</td>
<td>4.25</td>
<td>0.74</td>
<td>4.42</td>
<td>0.66</td>
</tr>
<tr>
<td>Trainer attitude towards trainees performing operative cases</td>
<td>4.58</td>
<td>0.74</td>
<td>4.80</td>
<td>0.48</td>
</tr>
<tr>
<td>Trainer to trainee feedback - quality</td>
<td>4.41</td>
<td>0.75</td>
<td>4.39</td>
<td>0.72</td>
</tr>
</tbody>
</table>

*Faculty trainers were not asked questions relating to supervision as these were felt to be more appropriate for trainees to answer.
6.2.4: Discussion

This chapter describes a regional quality improvement program for surgical training. The S-QAT developed in chapters 4 and 5 was used to assess and demonstrate variability in surgical training quality across a training region. Supervision of trainees and the balance between supervision and independence were identified as strengths of current training programs. Potential areas for improvement included the structured teaching programmes, training evaluation and quality control. Together with the structured teaching programmes, provision of designated training operating lists were identified both as potential areas for improvement and also as areas in which there was greatest variability in training quality. Multiple aspects of endoscopy training were revealed to demonstrate large variability in training quality. The S-QAT responses demonstrated that there was clear overlap between areas of training quality that demonstrated room for improvement and those that demonstrated greatest variability. This identification of variability can facilitate the dissemination of best practices to other training centers.

The objectives of this quality improvement program were to provide a formative assessment of surgical training, identify areas of best practice and variation in training centers and subsequently to use this information to make recommendations to improve the quality of surgical training in the North West London training region. The results revealed generally high standards of supervision in most aspects although notably the level of supervision in the outpatient clinic was not rated as highly as that in endoscopy, operating room or on the wards. The
free text quotes below highlight some of the differences between high and low quality supervision:

‘Out patient notes all vetted by consultant and appropriate mix of new and review patients allocated for trainee dependent on training level.’

‘Overcrowded very busy clinics with up to 70 patients, often have to sacrifice training theatre sessions to cover excess.’

The presence of training operating lists demonstrated the largest variation and was also the lowest rated element of surgical training quality as rated by both faculty trainers and trainee registrars. In the UK’s National Health Service training and service provision are generally provided simultaneously. This could potentially lead to imbalances between the two if either training or service provision were prioritised higher than the other as described in the free text quote below:

‘this would be ideal for training, but may not work for service delivery in context constrained resources’

In Aspiring to Excellence, the potential for these imbalances were explained. Owing to their objectively measurable outcomes, financial and service delivery targets are often better incentivised to an institution than resources for surgical training. Nevertheless, the fact that the
presence of training operating lists demonstrated high variability suggests that it can be done.

The next step would be to identify how this is being provided

‘Lists never specifically designed for training but workload and case mix is such that I feel the service lists are unintentionally excellent for training’

The free text quote above demonstrates that training may sometimes be inherently more suited to certain institutions. However, in other cases it may be simply adding some extra time per case, or careful selection of cases according to the seniority and competency of the trainee. Identification of these areas of best practice would allow this to be disseminated to the entire training region. Furthermore, if certain centres lend themselves to high quality operative training then identifying these could facilitate arrangement of blocks of targeted training. An example of an intervention to improve the provision of training operating lists can be found in one of Health Education England’s ‘Better training better care’ pilot projects: Increasing learning opportunities in surgery (HEE). In this pilot, a proportion of existing operating lists were identified and ‘ring-fenced’ as BTBC training operating lists for core surgical trainees (CSTs). This led to a 65% increase in the proportion of cases where a general surgery CST received supervised training.

Structured teaching programmes were the other area consistently being rated poorly and demonstrating large variation in rating. All forms of skills demonstrated high variability in quality rating, including technical, non-technical and research skills training. This is consistent
with the finding in chapter 3 that residents in a US training program received approximately three times as much teaching (2.3 hours per week versus 0.7 p <0.001) than those in London. They also had much better access to simulation facilities. North West London does have a regional skills teaching programme but based on the results of these S-QAT responses, the structure and organisation of this should be examined to optimise the benefits to the trainees.

Training evaluation itself and quality control measures for surgical training were highlighted as a potential area for improvement. Quality assessment is well established in the higher education sector. World university assessments have become an important aspect of the reputation of institutions and create large amounts of media interest (BBC 2014). Rankings have their limitations however, the transparency in the quality assessment methodology and the dynamic nature by which they evolve undoubtedly lead to resources being focused into improvement in any weak areas identified. The evaluation of quality in clinical outcomes within surgery has been a successful driver for quality improvement. In North America, the objective of the National Surgical Quality Improvement Program was to enhance the quality of surgical care through a comparative clinical outcomes assessment program (Khuri et al. 1998). There is evidence that the NSQIP program has been successful in its objective with proven improvements in quality of clinical care (Compoginis and Katz 2013). In this study evaluating vascular surgical site infections, NSQIP data assessment facilitated recommendations for changes to practice that led to a reduction in vascular surgical site infections. Similarly in the UK the introduction of clinical outcomes reporting through the Dr Foster unit (DFRL 2012) and the establishment of the Research Assessment Exercise (REF) have led to considerable investment in
resources to improve clinical and research quality. This chapter describes the first steps towards longitudinal assessment of surgical training quality which is essential for a full surgical training quality improvement program. The potential benefits could be to increase the focus on and resources to improve surgical training quality and to provide formative feedback and recommendations on how best to achieve this.

The main limitation of this study is the non-response bias. Although 93% of training centers responded, the response rate of Faculty trainers and trainee general surgery registrars was approximately half that at 47%. It is possible that those responding were the most engaged with training and hence may have rated training quality higher. In the absence of making completion of the S-QAT mandatory, the response rate will inevitably be less than 100% and the equivalent response rates for both trainers and trainees is encouraging. Free text feedback suggested an impression that the S-QAT was too long and would benefit from being shortened and the low completion rate would support this free text feedback. In particular, a number of Faculty trainers stated they felt some questions were more suitable for trainees rather than Faculty. Indeed four of the highest rated quality markers were related to Faculty trainers’ direct actions towards trainees; their attitude to trainees performing operative cases, their approachability, the rapport between the trainer and trainee and the level of independence afforded to the trainee. It is plausible that Faculty trainers will demonstrate some favourable bias to these quality markers. This is supported by the fact that in addition, all four markers feature in the lowest variability markers in table 6f. Therefore removing these from the Faculty
trainers’ questions will facilitate shortening the S-QAT with little negative effect on identifying variability in training quality.

6.2.5: Conclusion

This regional quality improvement initiative highlights variation in surgical training quality. Identification of variation can facilitate dissemination of best practices to other Training Centers. Greatest variation appears to be centred around the provision of training operating lists and the structure of teaching programmes. Centers should invest in protected teaching time in the operating room, more structured research and non-technical skills teaching programmes and place greater emphasis on training evaluation and quality control of training. The S-QAT tool has proven to be successful in demonstrating variation in training quality and providing data for useful formative feedback. Further refinement of the S-QAT should be dynamic, taking into account the feedback received in each iteration of its use. This will allow the S-QAT to become more streamlined and potentially encourage greater response and completion rates.
Chapter 7: Video Based Coaching to Improve the Quality of Surgical Training

The following chapter was published as:

7.1: Introduction

The focus of the thesis thus far has been to identify variations in surgical training quality and to develop a methodology to evaluate quality in surgical training. The final aim of this thesis was to develop and pilot interventions that can enhance the quality of surgical training. Improvements in surgical training tend to focus on increasing the quantity of training through either time or caseload. However, the introduction of work hour restrictions in particular, have made it essential to make surgical training more efficient. A hypothesis at the beginning of the thesis was that training improvements could be made without increasing the quantity of training but by improving the quality of training. These next two chapters will seek to develop, pilot and evaluate methods to pragmatically improve surgical training quality within the confines of current working conditions.

The Temple report recommended making ‘every moment count’ for trainees while they worked. It acknowledged how the introduction of the European Working Time Directive (EWTD) and the resultant move towards shift working, had led to reduced continuity between a trainee and their trainer. Simulators and simulation based training was supported in the Temple report (Temple 2010) as an innovative method to accelerate learning. However, despite increasing evidence of the benefits for simulation, its implementation has presented many challenges (Zevin et al. 2014). Limited resources in terms of both costs and also expertise in the use of simulation have been major obstacles to its widespread use. A potential alternative or adjunct to the use of simulation is the use of coaching. The Temple report supported the use of coaching and in the Shape of Training report (Greenaway 2013) the adoption of coaching and
mentoring techniques were cited as a key element of a successful apprenticeship-based arrangement. The sports industry employs video-analysis and coaching to maximize improvement from every performance (Wilson 2008). Whilst coaching is often associated with the sports industry (Jones and Kingston 2007), many professions and disciplines now utilize coaching to improve performance. Chapter 4 demonstrated the importance of trainer-trainee interaction and more specifically that continuity with the same trainer was an indicator of high quality training. Video based coaching could potentially offer a method to increase continuity between trainers and their trainees and maximising the learning from every clinical case performed.

Coaching is a process that aims to facilitate performance enhancement by identifying, focusing on, and achieving specific goals. In recent years, performance coaching has become known as an effective and widely accepted tool for professional development across many disciplines such as business (Deloitte), education (Van Niewerburgh 2012) and in healthcare (Gattellari et al. 2005). This emerging model of coaching is grounded in established adult learning and psychological concepts, which aims to foster self-directed learning, through performance evaluation and feedback, goal clarification, a collaborative approach to problem solving and action planning for the future (Cavanagh et al. 2005; Stober and Grant 2006). In contrast to instructive teaching or training methods, which deliver the knowledge and skills required to perform a task, coaching is a process where participants engage in a one-to-one reflective dialogue to monitor and evaluate progress towards their specific goals, and to modify action plans based on feedback (Cavanagh et al. 2005; Whitmore 1992). The ‘monitor-evaluate-modify’
element of this process establishes a cycle of self-regulated behaviour, which is a key process in creating intentional behaviour change (Carver and Scheier 2001; Laske 2004). Drawing an analogy between surgery and other disciplines such as business, it is possible that this form of collaborative performance review, found in one domain could be applicable to the other. A similar strategy has proven successful when importing other performance enhancement techniques such as mental practice (Arora et al. 2011) and deliberate practice (Crochet et al. 2011) from sporting and music disciplines and applying them to surgical skills performance. Coaching can facilitate surgeons to develop specific action plans to pursue lower-order targets and then to monitor and evaluate progression towards longer-term goals.

The coaching literature offers a number of structured and standardised methods and models, which can be adapted to a range of coaching contexts. The challenge for a surgeon is to find a coach who can be present while they are operating. With the widespread adoption of laparoscopic surgery, obtaining surgeons’ view video recordings of the operative field has become relatively straightforward. Coaching using videos of operative cases permits more flexibility and is a pragmatic solution to the challenges of obtaining real-time coaching. In this chapter, a coaching framework that can be applied to laparoscopic surgical skills performance is first identified. The coaching framework is then used to provide coaching with video recordings of procedures performed by the subject in a simulated setting. Laparoscopic skills performance is then compared between two randomised groups on real tissues using cadaveric porcine procedures. The objective of this study was to evaluate whether video-based coaching can enhance the quality of laparoscopic surgical skills performance.
7.2: Methods

7.2.1: Identification of a Coaching Framework for Laparoscopic Surgical Skills

Performance

A systematic, cross-disciplinary and multi-modal search was performed to identify and develop a suitable surgical coaching model. The majority of peer-reviewed empirical-research publications on coaching appeared in the behavioural sciences databases. There is consensus in the literature that the coaching process facilitates goal attainment by helping individuals to:

“(i) Identify desired outcomes/specific goals,
(ii) Enhance motivation by identifying strengths and building self-efficacy,
(iii) Identify resources and formulate specific action plans,
(iv) Monitor and evaluate progress towards goals, and
(v) Modify future actions. (Grant et al. 2010)”

Performance coaching is generally goal-oriented. The assumption is that the client (surgeon) knows what goals they want or need to achieve (e.g. reduction in operating time), or that they are capable of defining them with facilitative assistance. The GROW model (Alexander and Renshaw 2005; Landsberg 1997; Whitmore 1992) is an example of such a coaching methodology. The GROW model is a tool which can be used in many different coaching scenarios and is an effective technique often employed by both new and experienced career coaches. The GROW model offers a way of structuring coaching sessions to facilitate a balanced discussion. It can also be viewed as a framework around which future coaching techniques and
approaches can be developed. The GROW model guides a coaching conversation through four vital stages of goal-oriented coaching, namely Goals, Reality, Options and Wrap-up.

- Goals – focuses on specific targets that the client wishes to achieve.
- Reality – explores the true nature of the problem (performance review).
- Options – formulation of effective solutions, particularly to the issues that prevent the client from achieving their goals.
- Wrap Up – develops an action plan for the candidates to move towards their originally stated goals, examines potential obstacles and implementation strategy to overcome them.

Figure 7a depicts how the GROW model in combination with some key coaching questions was adapted for use in video based coaching.
Figure 7a: The 8 questions coaching framework for surgery developed with the GROW model (Alexander and Renshaw 2005; Landsberg 1997; Passmore 2010; Whitmore 1992).

- **Goal**
  - What is the purpose of the session?
  - What would be the most valuable topic of focus?

- **Reality**
  - What is the current situation? (Self assessment using global rating scales and video-based feedback from the coach)

- **Options**
  - What could you do to improve in those areas?
  - If you could only implement one of these options in the next procedure, what would it be?
  - Describe how you would implement this option?

- **Wrap-up**
  - What are the challenges to implementing this change and how to overcome them?
  - Action plan: What will you do, when?
Figure 7b: Flow of subjects through the study

20 novices recruited

Baseline skills test
LAP Mentor®

Trained on an evidence-based VR LAP Mentor® curriculum

Blocked randomisation into two groups (1:1 ratio)

Control Group (n=10)
5 LCs performed by each participant
30 minutes of an online surgical tutorial between each LC

Lost to follow-up (n=0)

Porcine LC x 2
Analysed (n=10)

Intervention Group (n=10)
5 LCs performed by each participant
30 minutes of video-coaching session with expert faculty between each LC

Lost to follow-up (n=0)

Porcine LC x 2
Analysed (n=10)
7.2.2: Subjects

The CONSORT diagram in Figure 7b demonstrates the study design. Laparoscopic novice subjects were medical students who voluntarily expressed an interest in study participation. Subjects were met by a study investigator and provided with written and verbal information before consent was obtained. Predefined exclusion criteria included subjects that had previously completed a laparoscopic skills training curriculum. The Institutional Review Board was contacted for ethical approval and a waiver was recommended.

7.2.3: Equipment & Setting

The study was conducted in a dedicated laparoscopic skills simulation centre at St Mary’s Hospital, Imperial College London. The Virtual Reality (VR) phase of the study employed the use of a VR laparoscopic simulator Lap Mentor® (Simbionix Corporation, Cleveland, OH). A previously validated VR laparoscopic cholecystectomy (LC) training curriculum was used to train the subjects to competency (Aggarwal et al. 2009). A high-definition SONY® digital camcorder was used to record all full procedures (including both VR and porcine LCs) performed after completion of training. In addition, FRAPS real time video capture software (Beepa Pty Ltd, Melbourne, Victoria, Australia) was used to record the simulator display. This provided a high-resolution exact replica of the subject’s operative view. Porcine laparoscopic cholecystectomies were recorded using a high-definition laparoscopic camera to capture laparoscopic views in addition to the camcorder recordings. A WD® external storage drive was used to store video data from all recording devices. Coaching sessions were delivered in a multimedia equipped private meeting room. VR LC video recordings displaying the operative field from the Lap
Mentor® were displayed on a fifty inch LCD monitor. The digital camcorder recordings were displayed on a twenty-three inch LCD monitor.

### 7.2.4: Baseline testing

The Lap Mentor® VR simulator was employed for baseline laparoscopic skills assessments of all subjects. Instructions were provided for two construct validated laparoscopic skills tasks; Task 1: ‘clipping and cutting’ and Task 2: ‘two-handed manoeuvres’ (Aggarwal et al. 2009). Subjects were permitted a single attempt of each task for practice. The subsequent attempt was recorded as their baseline performance.

### 7.2.5: Training to Competence

After completing baseline performance assessment, a standardised and evidence-based curriculum (Aggarwal et al. 2009) was followed to train all subjects to competently perform a laparoscopic cholecystectomy procedure. Training started with basic laparoscopic skills tasks. Subjects needed to achieve validated competence standards on two consecutive attempts in order to progress to the next set of tasks. These were procedural tasks that were once again repeated until validated competence standards were achieved on two consecutive attempts. In the final stage of the curriculum, subjects performed a full laparoscopic cholecystectomy. Again the required competence standards needed to be achieved on two consecutive procedures. Training sessions were limited to two hours each with no more than two sessions permitted in one day.
7.2.6: Randomisation

Randomisation took place after trainees completed the training curriculum to eliminate bias during the training phase. Therefore subjects were deemed competent to perform the simulated procedure prior to allocation to the control or intervention group. Blocked randomisation was employed to allocate equal numbers of subjects on a 1:1 ratio to either intervention or control groups. Random sequence blocks of 10 were generated through online randomisation software (Sealed-Envelope® Ltd., London, UK). Allocation concealment was maintained by concealing the randomisation sequence from study investigators involved in recruitment or training of subjects.

7.2.7: Control

Control subjects performed a standard VR LC five times with no feedback at any time, during or after the case. If the subject requested a change in the laparoscopic field of view, a trained laparoscopic assistant was available to move the camera. Upon completion of each repetition of the VR LC, subjects were shown thirty minutes of online surgical lectures. Subjects were made aware that comfort breaks of ten minutes were available upon request between cases.

7.2.8: Intervention: Video-Based Coaching

Intervention subjects performed a standard VR LC five times with no feedback given during or after the case until the coaching session. An experienced laparoscopic surgeon with greater
than six years of laparoscopic cholecystectomy experience provided coaching at the end of each completed VR LC in a multimedia equipped private meeting room as described above.

A maximum of thirty minutes was permitted for each coaching session. The sessions followed the 4 stage framework of the GROW model. Examination of the coaching literature highlighted some key questions which could be used to facilitate the coaching session while following the GROW model (Passmore 2010). The ten questions were modified for use with laparoscopic surgical skills training (Figure 7a). Two questions were removed as they were not relevant for laparoscopic skills coaching. In the Goals stage, subjects were asked to identify elements of their performance they felt they wanted to concentrate on. In the second stage of the GROW model; Reality, subjects used three validated global rating scales of intra-operative technical skills to conduct a self-evaluation of their performance in the previous VR LC.

- Global Operative Assessment of Laparoscopic Skills (GOALS) (Vassiliou et al. 2005).
- Objective Structured Assessment of Technical Skills (OSATS) (Martin et al. 1997).
- Operative Performance Rating System (OPRS) for LC (Larson et al. 2005).

Following self-evaluation using the global rating scales, the video recording of the previous VR LC was played back. Both the subject and coach were able to stop, rewind or fast-forward the footage as required. The coach focused on the elements of the performance identified in the goals stage to deliver constructive feedback to the subject. Subjects were encouraged to be interactive with their coach and direct the coaching towards areas they were most concerned with. In the Options stage, the coach discussed potential methods to improve the performance. Subjects were asked which coaching recommendations would be most useful for the next
procedure and what challenges they would face to implementation of these recommendations. Subjects were asked to describe their objectives for improvement in the performance of their next procedure in the final Wrap-up stage. They would identify how recommendations from the coaching session could facilitate successfully achieving these objectives. A study investigator moderated every coaching session between a coach and the subject and produced a three-point summary of the recommendations for the subject to take with them.

7.2.9: Porcine laparoscopic cholecystectomy

Following completion of the five VR LCs during the intervention phase, both control and intervention subjects performed two porcine laparoscopic cholecystectomies. The cadaveric porcine laparoscopy model has been previously described (Aggarwal et al. 2007) and served as a simulated measure of skills transfer from the VR setting to real tissue.

7.3: Outcome measures

Primary outcome was the quality of laparoscopic skills performance in the porcine LC procedures assessed using the following global rating scales of laparoscopic skill: GOALS (Vassiliou et al. 2005), OSATS (Martin et al. 1997), OPRS (Larson et al. 2005). A blinded trained surgeon assessor rated laparoscopic skills performance on the porcine LCs with the 3 global rating scales using laparoscopic video recordings of the procedures. A second trained surgeon assessor blindly reviewed a randomly generated twenty five percent sample of the
videos. The inter-rater reliability between both surgeon assessors was analysed as described below.

Secondary outcomes of time and dexterity metrics were assessed using the validated Imperial College Surgical Assessment Device (Bann et al. 2003). Laparoscopic skills performance was also measured during the VR LCs using:

- Global rating scales (GRSs) as described above to assess the quality of performance
- Validated simulator metrics including time, number of movements and path length of instruments (Aggarwal et al. 2009).

**7.4: Statistical analysis**

An a priori power analysis was performed in order to facilitate sample size selection. Similar studies from this research team revealed a thirty to thirty-five percent rise in the global rating scale scores following a deliberate practice training intervention (Crochet et al. 2011). Eight subjects or more were needed to demonstrate this effect with power = 0.80 and $\alpha = 0.05$. To compensate for any missing data or failure to complete the trial protocol, a decision was made to recruit ten subjects per group. The Statistical Package for the Social Science version 20.0 (SPSS, Chicago, IL), was used to conduct statistical analysis. Statistical significance was pre-defined as $p<0.05$ and results are reported as median values. A non-parametric statistical approach was used with the Mann–Whitney U test employed to analyse inter-group differences between unrelated samples. The Wilcoxon signed-rank test was employed to compare within group related samples. Inter-rater reliability of the blinded surgeon assessors was evaluated
using the Cronbach’s alpha statistical test. P-values are reported to 3 decimal places and the data is reported to 3 significant figures.

7.5: Results

7.5.1: Demographics

A total of twenty subjects were recruited; ten were allocated to the control group and ten to the intervention group. All subjects in both groups completed the study. Female:Male ratio was 3:7 in the control group and 1:9 in the intervention group. Median age was 20.5 years with a range of 19-22.5 years in the control group and 20 years with a range of 19-22.8 years in the intervention group. Completion of the VR LC curriculum took a median of 7 hours (range 6-10 hours) on 4 separate days (range 3 to 5 days).

7.5.2: Baseline laparoscopic skills

No significant differences were observed at baseline in the two validated laparoscopic skills task performances between intervention and control subjects; Task 1: ‘clipping and cutting’ – time (138 versus 122 seconds, P=0.668) and Task 2: ‘two-handed manoeuvres’ - time (105 versus 115, P=0.796), total number of movements (126 versus 157, P=0.159), and total path length (379 versus 415, P=0.315). In addition, performance in the first VR LC following the training phase again demonstrated no significant differences between interventions and controls in the quality of the procedure as assessed by global rating scales [GOALS:(13.5 versus. 13, P=0.471),
OSATS: (16.5 versus 16, P=0.324), OPRS: (26.5 versus 25, P=0.183). No significant differences were observed between intervention and controls during VR LC1 in the validated simulator-generated parameters of time taken (442 seconds versus 471, P=0.700), total number of movements (408 versus 435, P=1.00) and total path length (726 versus 688, P=0.739).

7.5.3: Virtual Reality laparoscopic cholecystectomies

Following VR LC1 that served as a baseline, for each subsequent repetition, intervention subjects significantly outperformed control subjects on all GRSs (Table 7a). These results are displayed graphically in Figure 7c using the OPRS rating scale. The global rating scores remained stable within the control group with no significant changes during the five virtual reality laparoscopic cholecystectomy repetitions. Observed inter-rater reliability between the two blinded, trained surgeon assessors was excellent for all global rating scales GOALS (Cronbach’s α = 0.941), OSATS (Cronbach’s α = 0.964) and OPRS (Cronbach’s α = 0.963).

Table 7b demonstrates the secondary outcomes, the VR simulator-generated metrics. Intervention subjects initially became slower and total path length and total number of movements increased. Intervention subjects’ performance plateaued at LC2 for time taken and total path-length, and LC3 for total number of movements. Control subjects initially became quicker and reduced total path-length and total number of movements. Control subjects’ performance plateaued at LC3 for time and path-length, and LC4 for movements.
Table 7a: Virtual Reality Laparoscopic Cholecystectomy Global Rating Scores (a) GOALS, (b) OSATS, (c) OPRS

(a) GOALS scores inter-group comparison

<table>
<thead>
<tr>
<th>LC1</th>
<th>LC2</th>
<th>LC3</th>
<th>LC4</th>
<th>LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) OSATS scores inter-group comparison

<table>
<thead>
<tr>
<th>LC1</th>
<th>LC2</th>
<th>LC3</th>
<th>LC4</th>
<th>LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) OPRS scores inter-group comparison

<table>
<thead>
<tr>
<th>LC1</th>
<th>LC2</th>
<th>LC3</th>
<th>LC4</th>
<th>LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GOALS scores intra-group comparison (p-value)

<table>
<thead>
<tr>
<th>Intervention</th>
<th>LC1 v LC5</th>
<th>LC2 v LC5</th>
<th>LC3 v LC5</th>
<th>LC4 v LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0.002</td>
<td>0.023</td>
<td>0.135</td>
<td>1</td>
</tr>
<tr>
<td>Control</td>
<td>0.656</td>
<td>0.563</td>
<td>0.617</td>
<td>0.531</td>
</tr>
</tbody>
</table>

OSATS scores intra-group comparison (p-value)

<table>
<thead>
<tr>
<th>Intervention</th>
<th>LC1 v LC5</th>
<th>LC2 v LC5</th>
<th>LC3 v LC5</th>
<th>LC4 v LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0.002</td>
<td>0.004</td>
<td>0.008</td>
<td>0.281</td>
</tr>
<tr>
<td>Control</td>
<td>0.344</td>
<td>0.273</td>
<td>0.344</td>
<td>0.938</td>
</tr>
</tbody>
</table>

OPRS scores intra-group comparison (p-value)

<table>
<thead>
<tr>
<th>Intervention</th>
<th>LC1 v LC5</th>
<th>LC2 v LC5</th>
<th>LC3 v LC5</th>
<th>LC4 v LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0.002</td>
<td>0.004</td>
<td>0.008</td>
<td>0.375</td>
</tr>
<tr>
<td>Control</td>
<td>0.422</td>
<td>0.574</td>
<td>0.410</td>
<td>0.273</td>
</tr>
</tbody>
</table>
Figure 7c: The quality of virtual reality laparoscopic cholecystectomy performance as assessed by OPRS
Table 7b: Virtual Reality Laparoscopic Cholecystectomy Simulator metrics (a) Total Time Taken, (b) Total Number of Movements, (c) Total Path Length

<table>
<thead>
<tr>
<th>(a) Total Time Taken (seconds) inter-group comparison</th>
<th>LC1</th>
<th>LC2</th>
<th>LC3</th>
<th>LC4</th>
<th>LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td>Median</td>
<td>421</td>
<td>338</td>
<td>404</td>
<td>426</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>244</td>
<td>345</td>
<td>495</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>IQR</td>
<td>135</td>
<td>91</td>
<td>190</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td><strong>Control</strong></td>
<td>Median</td>
<td>342</td>
<td>292</td>
<td>337</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>244</td>
<td>345</td>
<td>495</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>IQR</td>
<td>135</td>
<td>91</td>
<td>190</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td><strong>P-value</strong></td>
<td>0.700</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Total Number of Movements inter-group comparison</th>
<th>LC1</th>
<th>LC2</th>
<th>LC3</th>
<th>LC4</th>
<th>LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td>Median</td>
<td>408</td>
<td>435</td>
<td>558</td>
<td>392</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>244</td>
<td>354</td>
<td>426</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>IQR</td>
<td>114</td>
<td>97</td>
<td>91.2</td>
<td>96.8</td>
</tr>
<tr>
<td></td>
<td><strong>Control</strong></td>
<td>Median</td>
<td>408</td>
<td>435</td>
<td>558</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>244</td>
<td>354</td>
<td>426</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>IQR</td>
<td>114</td>
<td>97</td>
<td>91.2</td>
<td>96.8</td>
</tr>
<tr>
<td></td>
<td><strong>P-value</strong></td>
<td>1.000</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) Total Path Length (cm) inter-group comparison</th>
<th>LC1</th>
<th>LC2</th>
<th>LC3</th>
<th>LC4</th>
<th>LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td>Median</td>
<td>726</td>
<td>688</td>
<td>894</td>
<td>866</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>456</td>
<td>427</td>
<td>825</td>
<td>312</td>
</tr>
<tr>
<td></td>
<td>IQR</td>
<td>153</td>
<td>288</td>
<td>204</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td><strong>Control</strong></td>
<td>Median</td>
<td>726</td>
<td>688</td>
<td>894</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>456</td>
<td>427</td>
<td>825</td>
<td>312</td>
</tr>
<tr>
<td></td>
<td>IQR</td>
<td>153</td>
<td>288</td>
<td>204</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td><strong>P-value</strong></td>
<td>0.739</td>
<td>0.002</td>
<td>0.002</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Path Length (cm) intra-group comparison (P-value)</th>
<th>LC1 v LC5</th>
<th>LC2 v LC5</th>
<th>LC3 v LC5</th>
<th>LC4 v LC5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td>0.002</td>
<td>0.014</td>
<td>0.193</td>
<td>0.922</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>0.004</td>
<td>0.004</td>
<td>0.027</td>
<td>0.375</td>
</tr>
</tbody>
</table>

![Table 7b: Virtual Reality Laparoscopic Cholecystectomy Simulator metrics](image-url)
7.5.4: Porcine laparoscopic cholecystectomy

Intervention subjects outperformed control subjects in Porcine LC1 [GOALS: (20.5 versus. 15.5, \(P=0.011\), OSATS: (21.5 versus. 14.5, \(P=0.001\), OPRS: (26 versus. 19.5, \(P=0.001\)) and Porcine LC2 [GOALS: (28 versus. 17.5, \(P=0.005\), OSATS: (30 versus. 16.5, \(P<0.001\), OPRS: (36 versus. 21, \(P=0.004\)]. These results are displayed graphically in Figure 7d using the OPRS rating scale. Excellent inter-rater reliability was observed for GOALS (Cronbach’s \(\alpha = 0.956\), OSATS (Cronbach’s \(\alpha = 0.881\) and OPRS (Cronbach’s \(\alpha = 0.908\) global rating scales between the 2 blinded surgeon assessors.

Intervention subjects took significantly longer than controls in both porcine LC1 (2920 versus 2004 seconds, \(p=0.009\)) and LC2 (2297 versus 1683, \(p=0.003\)). No significant differences were observed between intervention and controls in total number of movements: porcine LC1 (1771 versus 1734, \(p=0.743\)) and LC2 (1859 versus 1696, \(p=0.696\)). No significant differences were observed between intervention and controls in total path-length: porcine LC1 (338 versus 403 centimetres, \(p=1.00\)) and LC2 (415 versus 335, \(p=0.762\)).
Figure 7d: The quality of porcine laparoscopic cholecystectomy as assessed by OPRS
7.6: Discussion

The use of professional coaching has become routine in many disciplines including sports (UKsport 2014), business (Passmore 2010) and large governmental organisations such as NASA (NASA 2014). The concept of coaching and its potential use with surgeons has stimulated much discussion (Gawande 2011). The objective of this chapter was to empirically investigate the effect of video-based coaching on the quality of laparoscopic surgical skills performance. Baseline skills assessment revealed no significant differences between intervention and control groups. Prior to testing the intervention, subjects were first trained on a validated virtual reality laparoscopic skills curriculum. This required subjects to pass competence standards throughout the curriculum before progressing to each stage. Upon completion of training to competence, subjects were randomised. The first virtual reality laparoscopic cholecystectomy (VR LC) revealed no significant differences between groups in either the quality of the performance as judged by the global rating scales, or the simulator based time and dexterity metrics. This demonstrated the integrity of the randomisation process. Despite equivalent exposure to practical laparoscopic skills training, video-based coaching enhanced the quality of laparoscopic surgical performance on all subsequent VR LCs as demonstrated by all GRSs. Skills transfer and retention were demonstrated by enhanced quality of laparoscopic performance on both cadaveric porcine LCs when compared to control subjects. At a time when surgical trainees’ exposure to operative cases comes at a premium, video-based coaching is a potentially feasible method of maximising performance enhancement from every clinical exposure.
The enhanced quality of laparoscopic performance came at the expense of increased time in the case of the cadaveric porcine LCs and at the expense of simulator derived metrics; time, number of movements and total path length, on the VR LCs. This finding is consistent with previous skills based studies (Arora et al. 2011; Crochet et al. 2011) where time and dexterity metrics either did not improve (Arora et al. 2011) or deteriorated (Crochet et al. 2011) in the intervention groups when compared to controls, despite statistically significant improvements in the quality of laparoscopic performance as assessed by validated global rating scales. Economy of motion is a characteristic of a surgical expert that is reflected in the global rating scales of performance. However, economy of motion is only relevant if the procedure is performed to the same standard. For example, if the cystic duct is dissected out and skeletonised completely by both an expert and a novice then one would expect the expert to have used fewer movements. However, in this study it appeared that intervention subjects’ focus shifted from speed, and simulator metrics, to a greater focus on the quality of the dissection with a resulting increase in movements during the progression along the learning curve.

The effects of video based coaching for experienced surgeons would be interesting to observe as coaching in other domains is often used to improve experts’ performance in addition to during the early years of acquiring that expertise. For instance, professional tennis players receive intensive coaching in order to attain and also to maintain their professional status on the world tour. A limitation of the methodology used in this chapter is the use of surgically naïve medical students as subjects. This was a pragmatic decision taken to overcome the
challenges of recruiting expert surgeons as subjects, in addition to finding a suitably experienced coach for those expert surgeons. While the concept of coaching expert surgeons has been introduced (Gawande 2011) and qualitatively investigated (Hu et al. 2012), this has not been empirically investigated. In order to explore the concept of using coaching to improve an already competent performance, all subjects in this study were first trained on a rigorous competency based laparoscopic surgical skills curriculum that required competencies to be met to progress through to each stage (Aggarwal et al. 2009). The training took a median of 4 days to complete. Following completion of the VR LC curriculum, the control group marginally improved their time and dexterity metrics while maintaining the quality of their laparoscopic skills performances on the subsequent VR LCs. This stability of skills performance confirmed that the training phase had been successful at achieving competence.

Current principles of surgical education target training to competence with an assumption that a competently trained surgeon will continue to improve with experience. However, Ericsson’s work has discussed the concept of deliberate practice, which proposes that purposeful practice will result in greater gains than practice alone (Ericsson 2006). The results presented in this chapter demonstrate this. Following completion of the VR LC curriculum, the control subjects failed to improve on the quality of their laparoscopic skills performances while the intervention group continued to improve the quality of their performance despite equivalent exposure to practice cases. Practice alone may maintain quality but the addition of coaching facilitated subjects to continue to improve the quality of their performance. This has implications for residency training programs that often rely on repetition of a procedure with little opportunity
for coaching. The study protocol did not give the control group the opportunity to view their videos in order to compare the video-based coaching intervention against what is current clinical practice. In current clinical practice, residents and surgeons who are deemed competent at a procedure rarely have the opportunity to regularly review videos of their operative procedures. This study was designed to evaluate whether video-based coaching could further improve performance beyond competence. Self-evaluation of videos may be an interesting area for future study, particularly in the search for ways to make the clinical implementation of video-based coaching more time-efficient. There is evidence that self-evaluation of videos can have benefits for novice learners (Farquharson et al. 2013) and it would be interesting to see if this would also be the case for learners who have already achieved competency in a task.

The main challenge to the clinical implementation of coaching with the modified GROW framework will be to identify faculty time to fit this intervention into current training programs. Ideally surgeons would receive real-time coaching during all surgical procedures they perform. However, in practice once a surgeon is competent to perform a procedure, continued teaching of that skill has to compete with the responsibilities and commitments of any potential coach. This chapter provides evidence that surgeons, could continue to improve their laparoscopic skills performances by applying the modified GROW model of coaching to review video recordings of procedures at a time that is mutually convenient with a more experienced coach. While the study was limited to simulated procedures, skills transfer to real tissues was demonstrated with the cadaveric porcine procedures. Clinical implementation of video-based coaching may require pragmatic modifications to the schedule described in this chapter.
Coaching could be a weekly or monthly session where junior surgeons meet the chief surgeon with videos of their cases. Certain generic goals could be set up that will be applicable to the majority of coaching sessions for a particular procedure and level of experience. However, junior surgeons should still be encouraged to identify more specific targets at the start of each coaching session in order to inform the coach and thus maximise the learning. It may be that self-evaluation of videos by junior surgeons may permit cutting the duration of the videos to only critical steps of procedures to reduce the time needed for coaching sessions. The self-evaluation or ‘Reality’ stage of the GROW coaching model could be conducted by junior surgeons either on their own or in a group with other junior surgeons thus reducing the time required with the senior surgical coach whose time is often the most limited. The effects of this delayed video-based coaching would warrant further investigation. Nevertheless, this study describes an exciting opportunity to improve training quality by maximising learning from operative cases performed when a coach cannot be present.

7.7: Conclusion

There are many external pressures on the surgical exposure to operative cases that trainees can receive. This chapter has described the development and pilot testing of an intervention to improve surgical training by focusing on the quality of training delivered. Despite equivalent exposure to practical laparoscopic skills training, video-based coaching enhanced the quality of laparoscopic surgical performance on both VR and Porcine LC although at the expense of increased time. Video based coaching is a potentially feasible method of improving surgical training quality through maximizing performance enhancement from every clinical exposure. It
could provide a pragmatic method to improve the continuity between consultant trainer and surgical trainees, which was identified as a key marker of high quality training in chapter 4.
Chapter 8: An Immersive and Intensive Simulation Week

“Boot Camp” to Improve the Quality of Surgical Training

The following chapter was published as:


Multiple government reports into medical education discussed in the introduction to this thesis raised concerns over the integration between the undergraduate and postgraduate medical education curricula. The e-consultation reported in *Aspiring to Excellence* (Tooke 2008) found divided opinion as to whether the newly introduced Foundation Curriculum was adequately integrated with the undergraduate medical curriculum. The Tooke report (Tooke 2008) encouraged the acquisition of more clinical experience under supervision towards the end of undergraduate medical training. The transition from medical school to foundation doctor was also discussed in *A High Quality Workforce* (DoH 2008a) and in point 87 of the *NHS Future Forum* report (DoH 2012). The *NHS Future Forum* suggested restoring the practice of workplace shadowing to facilitate the transition. This recommendation was eventually adopted and shadowing is now mandatory (DoH) however it remains susceptible to the natural variations in patient presentations. Concerns have been raised about whether four days of shadowing is sufficient to ensure patient safety standards can be maintained (House 2012). The *Shape of Training* report (Greenaway 2013) also confirmed the need to provide support for transitions during a medical career with particular focus on the transition form medical school to Foundation Year 1. The objective of this chapter was to develop, pilot and evaluate a method to improve surgical training quality at the interface between undergraduate medical education and postgraduate surgical training. Applying the knowledge gained from the previous chapters would allow development of a structured immersive simulation course that could allow controlled teaching in a safe environment.
8.1: Development of SIMweek – An Intensive & Immersive Simulation Week

“Boot Camp”

8.1.1: Introduction

Starting as a surgical foundation doctor can seem a daunting experience. These feelings are vindicated by evidence to suggest reduced efficiency and increased mortality in hospitals following the end of year turnover of doctors (Young et al. 2011). This phenomenon is referred to as the ‘July effect’ in the USA or the ‘August effect’ in the UK. This changeover is an example of ‘cohort turnover’ in which many doctors exit the workforce and are replaced by a similar number of new doctors at a single time point. The consequences of this include an acute loss of experience and disruption to operations e.g. delays from induction to new computer systems. Researchers into the ‘July effect’ have called for training to be administered before the foundation role is started and suggest simulation may play a vital role (Young et al. 2011). One intervention to facilitate the transition from medical school to foundation doctor is to ‘shadow’ the current Foundation Doctor.

In addition to the potential benefits for patients’ safety, focusing on the graduating medical student during their transition into clinical practice could alleviate some of the anxiety associated with this step increase in responsibility. This second point should not be underestimated given that on a weekly basis 1 in 7 surgical interns considers giving up their surgical career (Antiel et al. 2013).
In 2004 the American Surgical Association Blue Ribbon Committee recommended that ‘surgery departments should strive to make surgical clerkship and resident preparedness courses of the highest quality’ (Debas et al. 2005). This chapter first describes a framework to develop an intensive simulated week that will recreate experiences and situations that a Foundation Doctor will face in their first weeks of practice. The simulated environment can be controlled to promote learning in a safe environment and thus facilitate the transition from student to intern. Later in the chapter, the course is evaluated to assess its effectiveness.

8.1.2: Methods

This chapter first considers the end-users of the course and their associated learning needs including both technical and non-technical skills. Subsequently the resources required are discussed and a strategy for the organisation of a course and selection of teaching faculty is proposed. Finally the costs involved in running a course are considered, as this will undoubtedly have a significant effect on the uptake of any proposed training intervention.

In order to provide evidence-based recommendations for the framework proposed in this chapter, a systematic review of published literature was initially performed searching for any reports of the development, implementation or evaluation of a simulated skills course or ‘boot camp’ to prepare incoming surgical Foundation doctors or interns for their posts. PubMed and Medline databases were searched using combinations of the following key words ‘surg*’, ‘boot’, ‘camp’. The search included all surgical specialties and the last search was performed on
the 1st February 2014. Reports focusing on the use of novel simulators or simulation equipment as opposed to a simulated skills course were excluded, as were letters, comments and descriptive articles not reporting any original data. Courses designed for students who were not entering their internships were also excluded e.g. courses to prepare 3rd year students for the fourth year surgery clerkships or first year cardiac surgery residents who had already completed general surgery residency. Where multiple publications referred to the same course with some duplication of data, the most recent publication including the most recent and comprehensive data was used and the older publications excluded. Reference lists of all articles identified by the original search were hand searched to identify any further relevant articles. The search flow diagram is depicted in Figure 8a. Data was independently extracted by 2 researchers. Any discrepancies were resolved by revisiting the original study until consensus was achieved.
**Figure 8a: Flow Diagram of Search Strategy**

Records identified through database searching (n = 102)

Additional records identified through other sources (n = 11 from reference lists)

Records after duplicates removed (n = 60)

Records screened by title (n = 60)

Full-text articles assessed for eligibility (n = 36)

Studies included in qualitative synthesis (n = 18)

Records excluded:
- n = 4 – focus on simulator, not boot camp
- n = 6 – letters, conference abstracts, reviews not presenting original data
- n = 14 – not about surgical training

Full-text articles excluded, with reasons:
- n = 4 – descriptive, not presenting original data
- n = 1 – faculty motivation
- n = 3 – focus on simulator or a specific intervention within the boot camp
- n = 3 – multiple publications
- n = 6 – not starting as an intern
- n = 1 – online only
### Table 8a: Characteristics of studies in literature review

<table>
<thead>
<tr>
<th>Author</th>
<th>Subjects</th>
<th>Methodology</th>
<th>Specialty specific</th>
<th>Format</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antonoff et al</td>
<td>20 interns, 15 completed assessments</td>
<td>Distributed: 3 x 90 minute modules over 3 weeks</td>
<td>✓</td>
<td>Live animal or cadaveric tissue</td>
<td>Confidence: Yes, No baseline: Yes, Technical skills: Yes, Non-Technical skills: Yes, Knowledge: Yes</td>
</tr>
<tr>
<td>Boehler et al</td>
<td>32 students, 8 interns as control</td>
<td>Intensive: 1 month</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Brunt et al</td>
<td>31 students</td>
<td>Distributed: 7 x weekly 3 hour sessions</td>
<td>✓</td>
<td></td>
<td>Improved, Improved</td>
</tr>
<tr>
<td>Esterl et al</td>
<td>16 students</td>
<td>Intensive: 4 weeks</td>
<td>✓</td>
<td></td>
<td>Improved, No baseline</td>
</tr>
<tr>
<td>Fernandez et al</td>
<td>30 interns</td>
<td>Distributed: weekly 1 and 3 hour sessions for 9 weeks</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Heppel et al</td>
<td>20 interns</td>
<td>Intensive: 1 week</td>
<td>✓</td>
<td></td>
<td>Interns were not assessed but interns evaluated the course</td>
</tr>
<tr>
<td>Krajewski et al</td>
<td>18 interns</td>
<td>Distributed: 2 x 2.5 hour sessions per week for 2 months and 25 web-based self study modules</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Laack et al</td>
<td>12 students, 10 completed assessments, 28 controls, 22 completed assessments</td>
<td>Intensive: 1 week</td>
<td>✓</td>
<td></td>
<td>Improved, 89% rated intervention as aspect that best prepared them for internship</td>
</tr>
<tr>
<td>Lossing et al</td>
<td>28 interns</td>
<td>Distributed: 6 x 4 hours per week for 8 weeks</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Marshall et al</td>
<td>11 interns</td>
<td>Distributed: 3 x 75 minute sessions within one week</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Meier et al</td>
<td>17 interns, 21 completed assessments</td>
<td>Distributed: Web based and one day of simulation</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Okusanya et al</td>
<td>28 students, 9 controls</td>
<td>Intensive: 5 days</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Parent et al</td>
<td>15 interns, 13 controls</td>
<td>Intensive: 2.5 days</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Peyre et al</td>
<td>6 interns, 23 interns as controls</td>
<td>Intensive: 3 weeks</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Sellén et al</td>
<td>186 interns</td>
<td>Intensive: 1 weekend</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Sonnadara et al</td>
<td>6 interns, 6 junior and 6 senior residents as controls</td>
<td>Intensive: 1 month</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Tacito et al</td>
<td>23 students</td>
<td>Intensive: 5 days</td>
<td>✓</td>
<td></td>
<td>Improved</td>
</tr>
<tr>
<td>Todd et al</td>
<td>8 interns</td>
<td>Distributed: 15 x 1 hour interactive sessions over 1 month</td>
<td>✓</td>
<td></td>
<td>General Surgery</td>
</tr>
</tbody>
</table>

**Methodology:**
- Distributed
- Intensive
- Distributed: weekly
- Distributed: 2 x 2.5 hour
- Distributed: 6 x 4 hours per week
- Distributed: 3 x 75 minute sessions
- Distributed: Web based
- Distributed: one day of simulation
- Distributed: 3 x 90 minute modules
- Distributed: weekly 3 hour sessions
- Distributed: 7 x weekly 3 hour sessions
- Distributed: 3 week
- Distributed: 3 months
- Distributed: 1 month
- Distributed: 11 x 1 hour interactive sessions
- Distributed: 3 months

**Assessment:**
- Confidence: Yes, No baseline: Yes
- Technical skills: Yes, Non-Technical skills: Yes, Knowledge: Yes
- Improved perceptions of performance form faculty and nursing staff but no objective improvement compared to controls on structured clinical skills assessment
- Improved perceptions of performance form faculty and nursing staff but no objective improvement compared to controls on structured clinical skills assessment
- Improved perceptions of performance form faculty and nursing staff but no objective improvement compared to controls on structured clinical skills assessment
- Improved perceptions of performance form faculty and nursing staff but no objective improvement compared to controls on structured clinical skills assessment

**Format:**
- Hands on procedures or scenarios
- Live animal or cadaveric tissue
- No baseline
- 98% rated intervention as aspect that best prepared them for internship
- General Surgery
- No baseline
- Improved
- General Surgery
- Improved
- General Surgery
- Improved
- Improved
- Improved
- Improved
- Improved
- Improved
8.1.3: Results

8.1.3.1: Systematic review of published literature

A total of eighteen relevant articles were identified and their characteristics are summarised in Table 8a. A number of articles positively described courses targeting medical students in preparation for their surgical internship posts (Boehler et al. 2004; Esterl et al. 2006; Laack et al. 2010; Okusanya et al. 2012; Peyre et al. 2006). The primary outcome of these studies was often self-assessed confidence (Esterl et al. 2006; Laack et al. 2010; Okusanya et al. 2012; Peyre et al. 2006; Todd et al. 2011). The ‘hands-on’ components of internship preparation courses were identified by students as being the most valuable (Boehler et al. 2004; Esterl et al. 2006; Selden et al. 2012). In particular, mock pages and clinical scenarios were identified as being useful (Boehler et al. 2004; Esterl et al. 2006; Marshall et al. 2000; Meier et al. 2005) and didactic lectures were identified as the weakest parts of the course (Esterl et al. 2006). Some studies demonstrated evidence of benefits to technical skills (Brunt et al. 2008; Krajewski et al. 2013; Parent et al. 2010; Sonnadara et al. 2012) and knowledge (Antonoff et al. 2009; Boehler et al. 2004; Fernandez et al. 2012; Selden et al. 2012; Tocco et al. 2013).

Most articles described surgical skills courses that were not assigned to a particular subspecialty. However, specialty focused ‘boot camps’ in orthopaedics and neurosurgery, have also demonstrated evidence of improved technical skills (Sonnadara et al. 2011) and knowledge (Selden et al. 2012). A number of courses used live animal labs (Brunt et al. 2008; Heppell et al. 1995; Lossing et al. 1992; Peyre et al. 2006) or cadaveric tissue (Boehler et al. 2004; Esterl et al. 2006; Peyre et al. 2006; Selden et al. 2012; Sonnadara et al. 2012; Tocco et al.
2013) to augment the course. However, in Tocco et al.’s operative anatomy course, they reported no additional benefit to students from performing cadaveric procedures where adequate non-cadaveric simulation models were available for the same procedure (Tocco et al. 2013). A longitudinal study of the successful implementation of a simulation ‘boot camp’ over 4 years (Fernandez et al. 2012) suggested that further expansion of the course was limited by the need to comply with duty hour restrictions. The authors suggested one solution may be to consider moving the training to the later undergraduate medical curriculum.

8.1.3.2: Identifying the end-user

The end users of the proposed course would be final year medical students who have graduated from medical school, about to embark on their first post as a Foundation Doctor. Competency in terms of knowledge and to some extent some procedural skills will already have been tested as part of the undergraduate examinations. Therefore the course needs to be tailored towards applying knowledge and skills to practical situations, rather than teaching new knowledge or skills. The competencies required in the first week as a Foundation Doctor are generic, therefore the course can be designed as a generic surgical course for all surgical (sub)specialties. A balance has to be struck between keeping the course generic to all specialties and being specific enough to be tailored to a sub-specialty where necessary.

8.1.3.3: Identifying the learning needs

A needs analysis was performed based on the literature review, UK Foundation Programme curriculum (FP 2012) and discussion of typical postgraduate year 1 surgical cases and scenarios.
The practical components of courses were identified by trainees as being particularly valuable (Esterl et al. 2006). Most courses identified targeted generic surgical skills but a small number of ‘boot camps’ describing courses tailored to surgical subspecialties were found in the literature review (Selden et al. 2012; Sonnadara et al. 2011; Todd et al. 2011). The tasks need to focus on the basic skills required in the first few weeks as a Foundation Doctor. There is no benefit in teaching advanced laparoscopic skills to a student who is more concerned with how they will independently place a urethral catheter using the correct aseptic technique.

i) Technical skills training

- **Generic** – These procedures were identified from the UK Foundation Programme Curriculum core procedures (FP 2011, 2012).
  - Venepuncture
  - Intravenous cannulation
  - Urethral catheterisation
    - Male
    - Female
  - Arterial Blood Gas sampling

Other procedures can be added to encourage students to aim beyond competency and to challenge them e.g. placing a nasogastric tube or placement of a central venous catheter.

- **Surgical** – These procedures have been developed from previous research into assessment of technical ability in surgical trainees (Mackay et al. 2003), Foundation
Programme Curriculum core procedures (FP 2011) and previous research into introduction of novices to the operating room (Patel et al. 2012).

- Administration of a local anaesthetic
- Suturing
  - Simple interrupted
  - Mattress sutures
- Excision of skin lesions
- Basic laparoscopic skills
  - Laparoscopic camera manipulation
    - 0 degree
    - 30 degree
- Introduction to an operating room environment
- Scrubbing and gowning

ii) Non-technical skills training

These are largely generic professional skills. The end-users are medical graduates, therefore knowledge has previously been tested in medical school examinations. In previous courses, didactic lectures have been identified as a weaker part of the course (Boehler et al. 2004; Esterl et al. 2006). The focus should therefore be to shift away from didactic teaching of knowledge and to learn how to manage the knowledge already acquired through experience of clinical scenarios.
Mock pages have been reported to be an effective way to do this (Boehler et al. 2004; Esterl et al. 2006; Marshall et al. 2000). Some common mock pages have been suggested in Table 8b. These were developed from discussions with current and past Foundation Doctors, previously reported mock pages (Todd et al. 2011) and the UK Foundation Programme curriculum (FP 2012) which outlines competencies for the recognition and management of the acutely ill patient.

The immersive environment will prepare the intern for the real world scenario where history and examination findings may be ambiguous. For instance, post-operative chest pain may be cardiac in origin or a sign of a pulmonary embolism or a postoperative chest infection. Students can be taught a strategy to make a full assessment including basic investigations and know when the time is right to call either for more detailed investigations and/or call for senior assistance. All this needs to be performed while maintaining control over time and prioritising tasks including the accurate documentation of all clinical events.
Table 8b: Common Mock Pages and relevance to Foundation Programme competencies

<table>
<thead>
<tr>
<th>Mock Page scenario</th>
<th>Foundation Programme Competency – Recognition &amp; Management of the Acutely Ill Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promptly assess acutely ill, collapsed or unconscious patient</td>
<td></td>
</tr>
<tr>
<td>Responds to acutely abnormal physiology</td>
<td></td>
</tr>
<tr>
<td>Manages patients with impaired consciousness, including seizures</td>
<td></td>
</tr>
<tr>
<td>Manages pain</td>
<td></td>
</tr>
<tr>
<td>Manages sepsis</td>
<td></td>
</tr>
<tr>
<td>Manages acute mental disorder and self-harm</td>
<td></td>
</tr>
<tr>
<td>Chest Pain</td>
<td>✓</td>
</tr>
<tr>
<td>Poor urine output</td>
<td>✓</td>
</tr>
<tr>
<td>Patient has fallen on the ward</td>
<td>✓</td>
</tr>
<tr>
<td>Post-operative pyrexia</td>
<td>✓</td>
</tr>
<tr>
<td>Post-operative pain</td>
<td></td>
</tr>
<tr>
<td>Hypotension/Fluid Management</td>
<td>✓</td>
</tr>
<tr>
<td>Acutely confused/distressed patient</td>
<td>✓</td>
</tr>
<tr>
<td>Management of electrolyte disturbances</td>
<td>✓</td>
</tr>
</tbody>
</table>
8.1.3.4: Identification of a suitable venue and resources

The current UK model of preparing for the Foundation Year by ‘shadowing’ relies on the natural but variable chance encounters with patients in the clinical setting. In addition, whilst immersive and intensive, real world clinical scenarios do not permit independent experience for the subject – they require supervision and are therefore unable to ‘learn from their mistakes’ since this would be ethically unacceptable. The simulated environment facilitates the teaching of independent thinking in a controlled and reproducible environment while maintaining patient safety standards.

- **Immersive simulation facilities** - The Simulated Operating Suite offers a fully equipped minimally invasive operating theatre with basic general surgical open and laparoscopic instruments (Aggarwal et al. 2004). This is shown in Figure 8b(i). The recently validated Simulated Ward is a fully equipped simulated ward with medically trained patient actors (Puchet al. 2013). This is shown in Figure 8b(ii).

- **Clinical Skills Laboratory** - Simulated clinical skills scenarios incorporating technical skills simulation described above including; venepuncture/cannulation arm, nasogastric tube insertion model, male and female pelvis models for urinary catheterisation and suturing equipment including sutures and skin pads.

- **Virtual Reality Simulators** - Laparoscopic skills simulators including validated laparoscopic camera manipulation tasks (Aggarwal et al. 2009).

- **Multimedia resources** - As the focus of the course is shifted away from didactic teaching towards hands-on and immersive training, only limited audio-visual facilities are
required for lectures. Basic audio-visual facilities are needed to display blood results and radiological imaging. A desktop or laptop computer should suffice if teaching is largely in small groups. There is the possibility that this could be replaced by ipad ‘apps’ (Pixmeo-SARL) in the future. The audio-visual facilities in the simulated operating suite and the simulated ward permit audio enhancement to improve the face validity of the environment in addition to recording facilities for potential debriefing.
Figure 8b: (i) Simulated Operating Suite (ii) Simulated Ward
8.1.3.5: Organisation of the SIMweek

As one of the objectives of the course is to simulate the real world environment it is important to incorporate experience of working late into the evening or even overnight. A proposed timetable would therefore include a combination of 3 standard shifts from 08:00 to 20:00 and one overnight shift from 20:00 to 08:00. Shifts would follow a regular structure that targets interactive scenarios and practical or technical skills in preference to passive didactic lectures. Based on the needs analysis, laparoscopic teaching should focus on basic skills such as camera navigation rather than intermediate laparoscopic skills. Repetition of commonly performed tasks and procedures will further enhance the realism of the immersive simulation experience. In addition, while the day should be structured, spontaneous mock bleeps throughout the day or night shifts and including during breaks will again enhance the realism. The subjects should be blinded as to the timing and nature of these mock bleeps. Table 8c describes a sample timetable.

8.1.3.6: Assessment of surgical practice

In order to provide formative feedback for the subjects and also for the course developers, an objective assessment process would be beneficial. This process should assess both objective competency in a range of clinical skills and in addition should include some indication of the subjects’ perceived confidence that they are ready to start as a surgical intern. Objective clinical skills assessments can be conducted using Objective Structured Clinical Examination (OSCE) stations.
### Table 8c: Sample Timetable

<table>
<thead>
<tr>
<th>Day</th>
<th>Time</th>
<th>Clinical Skills Laboratory</th>
<th>Simulated Operating Suite</th>
<th>Simulated Ward</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 AM</td>
<td></td>
<td>Blood Sampling Interpretation – (High K⁺) Female Urethral Catheterisation Arterial Blood Gas Sampling Interpretation (Hypoxia)</td>
<td>Surgical Instrument identification Scrubbing for Surgery Gloving and Gowning Laparoscopic Camera Navigation</td>
<td>Ward Round Mock Bleep Chest Pain Rewriting Drug Chart (with errors to be corrected)</td>
</tr>
<tr>
<td>1 PM</td>
<td></td>
<td>Suturing – Interrupted wound closure</td>
<td></td>
<td>Mock Bleep Post op pyrexia</td>
</tr>
<tr>
<td>1 Night</td>
<td></td>
<td>Rest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 AM</td>
<td></td>
<td>Local Anaesthesia administration Excision biopsy</td>
<td>X-ray interpretation – Bowel obstruction Surgical instrument identification</td>
<td>Ward Round Mock Bleep GI bleeding</td>
</tr>
<tr>
<td>2 Night</td>
<td></td>
<td>Rest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 AM</td>
<td></td>
<td>Male Urethral Catheterisation Blood Sampling Interpretation (Low Na⁺) X-ray Interpretation – NG placement Arterial Blood Gas Sampling Interpretation (Hypoxia with raised CO₂)</td>
<td></td>
<td>Ward Round Mock Bleep Poor urine output (retention)</td>
</tr>
<tr>
<td>3 Night</td>
<td></td>
<td>Rest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 AM</td>
<td></td>
<td>Rest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 PM</td>
<td></td>
<td>Rest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Night</td>
<td></td>
<td>Male Urethral Catheterisation Arterial Blood Gas Sampling Interpretation (Metabolic acidosis) Blood Sampling Interpretation (Low K⁺) Digital Rectal Examination</td>
<td></td>
<td>Mock Bleep Post op pyrexia Poor urine output Post op-pain Patient fall on ward Chest Pain Fluid management X-ray interpretation – Pneumoperitoneum #Neck of Femur</td>
</tr>
<tr>
<td>5 AM</td>
<td></td>
<td></td>
<td></td>
<td>Handover Post Take Ward Round</td>
</tr>
<tr>
<td>5 PM</td>
<td></td>
<td>Rest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Night</td>
<td></td>
<td>Rest</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8.1.3.7: Selection and Organisation of Faculty

There is a role for experienced faculty, ideally surgical consultant level. There is also a role for faculty who are currently in the role of Foundation Doctor (PGY1) to allow familiarity with the role being simulated. Recent data from the Toronto Orthopaedic Boot Camp team provide encouraging results for student-led learning although they highlight the difference between this and self-taught models (Sonnadara et al. 2013). Student-led learning would still require faculty supervision as a source of knowledge although there is the possibility that total faculty time could be reduced. There is a potential role for non-medical faculty in the future. Clinical skills may be taught by trained, non-medical faculty at medical school (Oswald et al. 2011). This may help to assist in continuity of the faculty, as they could be permanent faculty rather than rotating surgical trainees and consultants. Often senior nursing staff will have many years of experience in helping junior doctors in the transitional weeks from medical student to Foundation Doctor so they could provide useful feedback to subjects, in particular during the mock bleep scenarios. Subjects can be organised into pairs in order to allow 1:2, Faculty:Subject supervision ratios. However, feedback from subjects of the pilot course described later in this chapter suggested that ratios of up to 1:5 may be feasible without reducing the quality of the experience. Previously reported courses have used 1:4, Faculty to Subject supervision ratios with encouraging results (Okusanya et al. 2012).

8.1.3.8: Costs

A typical academic surgical residency program in the USA may have approximately 40 surgical interns per year including all surgical specialties. Excluding faculty expenses, the costs for the
program would amount to under $20,000 (or $455 per subject) to cover consumables, and venue/equipment costs. Table 8d gives a breakdown of the costs. The need for faculty and any potential reimbursement needs to be considered. Exact numbers and hours of faculty time required will depend on the desired faculty to student ratios and amount of administration that could potentially be shared with other departments. If running the SIMweek with a faculty to student ratio of 1:2, 10 students could be accommodated per week, requiring a minimum of 3 faculty per day and 1-2 faculty per night. Increasing the faculty to student ratio would help to keep costs down. For instance using a program with 40 interns as an example, if a faculty:student ratio of 1:4 were used then all students could be put through the SIMweek in two weeks thus requiring a total of 480 hours of faculty time. This could be met by eight faculty members each providing sixty hours of teaching time. Another method to reduce faculty costs would be to use trainees with an interest in teaching or potentially even trained non-medical instructors. Indeed there is evidence that non-clinicians can be trained to assess non-technical skills in surgery (Russ et al. 2012). In addition to teaching faculty, an administrator and a technician would be valuable assets as their costs would likely be offset by efficiency savings in the running of the course and the subsequent reduction of the more expensive faculty time required.
Table 8d: Costs for the SIMweek program using a typical Academic Surgical Residency program with 40 interns covering all surgical specialties as an example

<table>
<thead>
<tr>
<th>Item</th>
<th>No. needed per subject</th>
<th>No. of subjects</th>
<th>Total no. needed</th>
<th>Unit price ($)</th>
<th>No. in each unit (e.g. 'box of 5')</th>
<th>Price per item ($)</th>
<th>TOTAL ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABG syringe</td>
<td>3</td>
<td>40</td>
<td>120</td>
<td>212.81</td>
<td>100</td>
<td>2.13</td>
<td>255.38</td>
</tr>
<tr>
<td>Blood bottles</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>11.30</td>
<td>100</td>
<td>0.11</td>
<td>9.04</td>
</tr>
<tr>
<td>IV Cannulae</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>58.44</td>
<td>50</td>
<td>1.17</td>
<td>93.50</td>
</tr>
<tr>
<td>Vacutainer system</td>
<td>4</td>
<td>40</td>
<td>160</td>
<td>19.55</td>
<td>25</td>
<td>0.78</td>
<td>125.10</td>
</tr>
<tr>
<td>Catheter</td>
<td>4</td>
<td>40</td>
<td>160</td>
<td>16.25</td>
<td>5</td>
<td>3.25</td>
<td>519.98</td>
</tr>
<tr>
<td>Catheter pack (with gloves)</td>
<td>4</td>
<td>40</td>
<td>160</td>
<td>0.74</td>
<td>1</td>
<td>0.74</td>
<td>118.06</td>
</tr>
<tr>
<td>Instillagel</td>
<td>4</td>
<td>40</td>
<td>160</td>
<td>18.24</td>
<td>10</td>
<td>1.82</td>
<td>291.89</td>
</tr>
<tr>
<td>Gloves</td>
<td>7</td>
<td>40</td>
<td>280</td>
<td>6.80</td>
<td>100</td>
<td>0.07</td>
<td>19.03</td>
</tr>
<tr>
<td>Apron</td>
<td>3</td>
<td>40</td>
<td>120</td>
<td>10.24</td>
<td>200</td>
<td>0.05</td>
<td>6.14</td>
</tr>
<tr>
<td>Disposable tourniquets</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>3.88</td>
<td>25</td>
<td>0.16</td>
<td>12.41</td>
</tr>
<tr>
<td>Sterets</td>
<td>8</td>
<td>40</td>
<td>320</td>
<td>1.52</td>
<td>200</td>
<td>0.01</td>
<td>2.44</td>
</tr>
<tr>
<td>Guaze</td>
<td>7</td>
<td>40</td>
<td>280</td>
<td>4.71</td>
<td>100</td>
<td>0.05</td>
<td>13.19</td>
</tr>
<tr>
<td>LA equipment (needle &amp; syringe)</td>
<td>1</td>
<td>40</td>
<td>40</td>
<td>7.13</td>
<td>100</td>
<td>0.07</td>
<td>2.85</td>
</tr>
<tr>
<td>NG tube</td>
<td>1</td>
<td>40</td>
<td>40</td>
<td>15.01</td>
<td>10</td>
<td>1.50</td>
<td>60.04</td>
</tr>
<tr>
<td>Sterile gowns</td>
<td>3</td>
<td>40</td>
<td>120</td>
<td>67.98</td>
<td>30</td>
<td>2.27</td>
<td>271.92</td>
</tr>
<tr>
<td>Sterile gloves</td>
<td>3</td>
<td>40</td>
<td>120</td>
<td>32.70</td>
<td>50</td>
<td>0.65</td>
<td>78.49</td>
</tr>
<tr>
<td>Sutures</td>
<td>8</td>
<td>40</td>
<td>320</td>
<td>72.74</td>
<td>36</td>
<td>2.02</td>
<td>646.56</td>
</tr>
<tr>
<td>Skin pads</td>
<td>3</td>
<td>40</td>
<td>120</td>
<td>41.45</td>
<td>12</td>
<td>3.45</td>
<td>414.48</td>
</tr>
<tr>
<td>Actors (1 hour = 1 unit) ^c</td>
<td>6</td>
<td>40</td>
<td>240</td>
<td>39.25</td>
<td>1</td>
<td>39.25</td>
<td>9420.00</td>
</tr>
<tr>
<td>Hire of Simulated Operating Suite</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>6280.00</td>
<td>1</td>
<td>6280.00</td>
<td>6280.00</td>
</tr>
<tr>
<td>TOTAL for program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$18640.51</td>
</tr>
<tr>
<td>TOTAL per subject</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$466.01</td>
</tr>
</tbody>
</table>

^a Please note that faculty time has not been costed for in this table.

^b Rounded up to 2 decimal places. The exact figure was used when calculating the total cost.

^c Calculated as 3 day ward rounds plus one night ward round for each subject = 4 units. Three actors per ward round: 3x4 = 12 units. Two subjects per ward round: 12/2 = 6 units per subject.
8.1.4: Conclusion

There is evidence of increased mortality and reduced efficiency in hospitals due to the changeover of doctors at the end of the year (Young et al. 2011). This has been described as the ‘July effect’ in the USA and the August effect in the UK. Patient mortality was found to be increased during the August changeover at the same time as graduating medical students begin their clinical careers as interns (Jen et al. 2009). In addition to the potential benefits for patients’ safety, focusing on the graduating medical student during their transition into clinical practice could alleviate some of the anxiety associated with this step increase in responsibility.

While compulsory shadowing is one method to address the challenge, 4 days of compulsory shadowing is likely not enough on its own to protect patients (House 2012). This chapter proposes a framework for the development, organisation and implementation of an intensive simulation course to prepare graduating medical students for their role as junior surgical intern. Whilst the costs of the course have been described, the true cost effectiveness will be dependant on the benefits provided by the course. The next step is to evaluate the educational benefits to the surgical trainees and potentially the clinical benefits of such a course to the patients they look after.
8.2: Pilot Testing of An Intensive and Immersive Simulation Week “Boot Camp”
to Improve the Quality of Surgical Training

8.2.1: Introduction

Competence in medical school examinations assesses both cognitive and to some extent practical skills. The focus of these examinations is on testing knowledge and targeted technical or practical skills. However, this does not necessarily provide adequate preparation for the real world situations that a surgical intern will face. The purpose of this study is to pilot and evaluate an immersive and intensive simulated one-week course which has been described in the first part of this chapter. The course will recreate experiences and situations that a surgical doctor is likely to face in their first weeks as a clinician. It aims specifically to prepare final year medical students for the transition into their professional role as a junior doctor. This will allow the subject to learn and practise in a controlled simulated environment away from the patient. The hypothesis is that attendance at a one-week intensive simulation course will better prepare medical students for the transition to the role of Foundation Doctor as measured by:

- Primary Outcome: Clinical performance
- Secondary Outcome: Trainee confidence and clinical procedures logbook
8.2.2: Methods

8.2.2.1: Subjects

Final year students who had completed medical school and were due to commence their Foundation Year 1 placement in August 2012 were eligible for the study. This study population was chosen in part for pragmatic reasons, as recruitment of students would avoid the requirement for study or annual leave to be arranged with NHS employers if they were already practising clinicians. Furthermore it was felt that students about to commence their first clinical placement as a qualified doctor, would derive the most benefit from this boot camp course. Eligible subjects who voluntarily expressed an interest in participation were met by the study investigator. Both written and verbal information were provided and written consent was then obtained for all subjects. Advice regarding ethical approval was sought from the Institutional Review Board and a waiver was recommended.

8.2.2.2: Settings & Scenarios

All assessments and interventions took place in a controlled, simulated environment.

- Simulated Ward
  - A recently validated, immersive high fidelity ward simulator in which realistic perioperative patient encounters can be recreated (Pucher et al. 2013).

- Simulated Operating Suite
An immersive high fidelity replica of an operating theatre, complete with an operating table, anaesthetic equipment and surgical scrubbing facilities (Aggarwal et al. 2004).

- Clinical Skills Laboratory
  - Complete with simulation equipment for basic clinical procedures such as venepuncture and intravenous cannulation, male and female urinary catheterisation, arterial blood gas sampling.

- Virtual Reality Skills Laboratory
  - Providing access to the LapMentor virtual reality laparoscopic simulator (Simbionix Corporation, Cleveland OH).

The SIMweek intervention consisted of a one-week intensive simulation ‘boot camp’ using the environments described above. The interdisciplinary faculty consisted of general surgeons, gynaecologists, and nursing staff, with experience ranging from a current Foundation Doctor, to Consultant level. The week comprised of 3 ‘day shifts’ from 8am to 8pm and one ‘night shift’ from 8pm to 8am. The format for each day was similar; focusing on practical and technical skills rather than didactic teaching. Further details of the activities during the week of simulation can be found in Table 8e that displays the SIMweek timetable.
Table 8e: SIMweek Timetable

<table>
<thead>
<tr>
<th>Time</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0700</td>
<td></td>
<td></td>
<td></td>
<td>Handover</td>
<td></td>
</tr>
<tr>
<td>0800</td>
<td>Ward round</td>
<td>X-ray interpretation:</td>
<td>Blood sampling, Arterial blood gas</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bowel obstruction:</td>
<td>sampling</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NG tube placement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0900</td>
<td>Blood sampling,Female urethral</td>
<td>Ward round</td>
<td></td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td></td>
<td>catheterisation,Arterial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>blood gas sampling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>Local Anaesthesia administration,</td>
<td>Ward round -Group C</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Excision biopsy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>Mock bleep chest pain, rewriting drug chart</td>
<td>Mock Bleep, poor urine output, male urethral catheterisation</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td>1200</td>
<td></td>
<td>Mock Bleep, GI bleeding</td>
<td>X-ray interpretation:</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>bowel obstruction:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NG tube placement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1300</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>1330</td>
<td>Scrubbing and Gloving and Gowning</td>
<td>Mock bleep, post-op chest pain</td>
<td>Suturing</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td>1400</td>
<td>Surgical instrument identification</td>
<td>Rewriting Drug Chart</td>
<td>Suturing</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td>1500</td>
<td>Laparoscopic camera navigation</td>
<td>Male urethral catheterisation</td>
<td></td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td>1600</td>
<td>Break</td>
<td>Break</td>
<td>Break</td>
<td>Break</td>
<td>Break</td>
</tr>
<tr>
<td>1630</td>
<td>Suturing</td>
<td>Scrubbing and Gloving and Gowning</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td>1700</td>
<td>Suturing</td>
<td>Laparoscopic camera navigation (in VR lab)</td>
<td>Surgical instrument identification</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td>Time</td>
<td>Activity Details</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1800</td>
<td>Mock bleep, post op pyrexia</td>
<td>Blood sampling</td>
<td>Laparoscopic camera navigation</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td>1900</td>
<td>Scrubbing and Gloving and Gowning</td>
<td>Rest</td>
<td>Rest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Ward round</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2100</td>
<td>Mock bleep – Chest pain, Arterial blood gas sampling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2300</td>
<td>Mock bleep post op pyrexia – Blood sampling, cultures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0000</td>
<td>Mock bleep poor urine output – male urethral catheterisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0200</td>
<td>Mock bleep, new admission, digital rectal examination, pneumoperitoneum</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0400</td>
<td>Mock bleep – fall on ward, x-ray interpretation #NOF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0600</td>
<td>Mock bleep – post operative pain, fluid management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0700</td>
<td>Handover</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The procedures and tutorials were chosen to represent those most likely to be performed by the subjects in their first weeks as an intern. Therefore a decision was made not to teach intermediate laparoscopic skills but rather focus on basic skills such as laparoscopic camera navigation. In addition, the core procedures (FP 2011) for Foundation Year 1 (intern year) competencies were used in conjunction with the experiences of the broad ranging faculty to maximise the clinical relevance of the immersive course. In the UK satisfactory completion of the intern year is mandatory to obtain full registration from the General Medical Council. In order to enhance the realism, certain tasks were repeated throughout the course and the weighting of how many repetitions were based on the frequency with which these tasks are performed in the clinical setting. In addition, between stations, subjects had breaks and mock bleeps (or pager calls) occurred sporadically throughout the day (including during breaks). Every subject received at least 2 mock bleeps during the day shifts and at least 3 overnight. The number varied depending on their performance. For example if performance was felt to be lacking in one aspect such as initial assessment, then a further mock bleep would be administered in order to provide an extra chance to practice and build on the feedback received from the previous mock bleep. An example is provided in Figure 8c.
**Figure 8c: Example Mock Bleep Scenario**

<table>
<thead>
<tr>
<th>Bleep scenario:</th>
<th>Ask the nurse to:</th>
<th>Actions to take:</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Doctor, Mr Jones has not passed urine for 4 hours&quot;</td>
<td>- Flush the catheter</td>
<td>- If not catheterised, implies urinary retention - ?cause</td>
</tr>
<tr>
<td></td>
<td>- If not flushing, replace catheter</td>
<td>- If catheterised, suggests dry - ?cause (drain output, underfilled, GI upset, bleed etc)</td>
</tr>
<tr>
<td></td>
<td>- If flushing, administer IV fluid bolus</td>
<td>- Ensure adequately hydrated (IVF)</td>
</tr>
<tr>
<td></td>
<td>- Urine dip</td>
<td>- Check UO after 2-4 hours</td>
</tr>
<tr>
<td></td>
<td>- If NOT catheterised, bladder scan and catheterise if &gt;500ml.</td>
<td>Urgency of r/v? – Medium to high</td>
</tr>
<tr>
<td></td>
<td>- Repeat obs. Monitor UO</td>
<td>- Currently on IV fluids or no – if so, what rate?</td>
</tr>
</tbody>
</table>

Questions they should ask:
- Is Mr Jones catheterised?

IF NOT CATHETERISED:
- Pain?
- Urge to urinate?
- Bladder scan?

IF CATHETERISED:
- Is the catheter flushing?
- What is the UO trend?

THEN:
- Other obs (?high HR, ?low BP, ?pyrexial)
- Positive or negative balance according to chart
- Currently on IV fluids or no – if so, what rate?
8.2.2.3: Study Design

In the interests of feasibility for recruitment, a non-randomised controlled trial design was employed (as the intervention group were required to participate in a week long course, this was likely to conflict with travel plans which are common at this stage of training). The study design is outlined in Figure 8d. All subjects underwent baseline assessment of technical skills and confidence prior to starting their Foundation Year 1 posts. The control group had no further intervention beyond standard medical school training. From 2012 this now included one week of intern shadowing.

The intervention group underwent an immersive and intensive one-week simulation course, followed by post intervention assessment of technical skills and confidence. The intervention group continued with the compulsory Foundation Doctor shadowing. Follow up assessments of technical skills and confidence were performed for both intervention and control groups at 1 month and 6 months into their Foundation Year 1. All subjects were in addition instructed to keep a clinical logbook of procedures performed for the first month and for a 1-week sampling period at 4 and 6 months. A 1-week sample was used as it was felt to be unfeasible to request subjects to reliably record all clinical procedures performed for a 6-month period.
Figure 8d: Study Design

**Enrollment**
- Assessed for eligibility (n = 19)
  - Excluded (n = 0)

**Baseline assessments**: Clinical OSCE, Surgical skills, Confidence questionnaire

**Allocation**
- Allocated to SIMweek intervention (n = 10)
  - Received standard medical school training including 1 week shadowing (n = 10)
  - **Repeat assessments** prior to starting as a Foundation Doctor
- Allocated to Control (n = 9)
  - Received standard medical school training including 1 week shadowing (n = 7)
  - Did not receive shadowing as unable to begin their Foundation Doctor post (n = 2)

**Follow-Up**
- Repeat assessments at 1 month and 6 months (n = 10)
  - Clinical logbook of procedures over 6 months
  - Lost to follow-up (n = 0)
- Repeat assessments at 1 month and 6 months (n = 6)
  - Clinical logbook of procedures over 6 months
  - Lost to follow-up (n = 1)

**Analysis**
- Analysed (n = 10)
  - Excluded from analysis (n = 0)
- Analysed (n = 6)
  - Excluded as did not start Foundation Doctor post (n = 2)
  - Excluded as did not attend assessments (n = 1)
8.2.2.4: Outcomes

Primary outcome: Clinical Performance

Assessed using clinical OSCEs, surgical technical skills and laparoscopic skills assessments at:

- Baseline
- Post SIMweek intervention (intervention group only)
- One month into Foundation Year 1
- 6 months into Foundation Year 1

Secondary outcomes: (i) Self assessed trainee confidence questionnaires

- At baseline, post SIMweek (intervention group only) and at 1 and 6 months into Foundation Year 1 as for primary outcomes.

(ii) Clinical Procedures Logbook

- At weeks 1, 2, 3, 4 and at months 4 and 6.

Assessment tools

OSCEs

Basic clinical skills were assessed using Objective Structured Clinical Examination (OSCE) stations. The assessments were all procedures taken from the core procedures(FP 2012) for Foundation Year 1 competencies. Score sheets were derived from the competency assessments(FP 2011) provided by the Foundation Programme. A sample score sheet is provided in Appendix D.

- Venepuncture
Male urethral catheterisation

- Intravenous cannulation

- Arterial Blood Gas Sampling

Surgery specific OSCEs comprised three surgical technical skills stations and one laparoscopic skills station assessed using previously validated tasks and metrics (Aggarwal et al. 2009; Mackay et al. 2003; Martin et al. 1997)

- Interrupted suture closure of a wound – assessed using the objective structured assessment of technical skills (OSATS)

- Mattress suture closure of a wound – OSATS

- Excision of a skin lesion - OSATS

- Laparoscopic camera navigation – assessed using time and accuracy

Surgical faculty trained for assessment of medical school final OSCE examinations performed all scoring. OSCE stations were video recorded (omitting the face to facilitate anonymity) and a 25% sample was assessed by a second blinded reviewer to ensure accuracy of scoring.

**Confidence Questionnaire**

This was assessed using a self-assessment questionnaire consisting of 35 questions relating to confidence, each marked on a 5-point Likert scale (the questions are provided in Appendix E). This was developed based on those used for the previously mentioned studies on surgical skills courses for medical students (Esterl et al. 2006; Okusanya et al. 2012; Peyre et al. 2006; Todd et al. 2011). Questions that were not felt to be relevant to UK Foundation Year doctors were removed.
Clinical Procedures Logbook

Telephone interview of subjects for the number of bedside procedures performed either independently or whilst supervised.

8.2.3: Statistical Analysis

Non-parametric statistical methods were employed with the use of Mann-Whitney U test for between group comparisons and the Wilcoxon signed-rank test to analyse data within groups. Inter-rater reliability for clinical OSCEs was assessed using Cohen’s kappa. Inter-rater reliability for surgical technical skills was assessed using Cronbach’s alpha. All statistical analyses were performed using the Statistical Package for the Social Science version 20.

8.2.4: Results

Ten students were recruited to and completed the intervention arm. Nine students were recruited as controls however two students were unable to start their Foundation Year 1 posts and were therefore excluded from all analyses. One control was lost to follow up after 3 weeks and was therefore also excluded from all analyses, leaving six remaining students in the control arm.
8.2.4.1: Baseline

No significant differences were demonstrated in baseline clinical OSCE, surgical technical skills, confidence and laparoscopic camera operating scores between intervention and control groups.

8.2.4.2: Primary Outcome: Clinical Performance

Table 8f and Figure 8e display the results for the basic clinical OSCEs and surgical technical skills assessments. The intervention group outperformed the control group at both 1 and 6 months in both the clinical OSCEs and the surgical technical skills. Intervention performance significantly improved from baseline to post-course, however no significant improvements were made in performance from post-course to 1 month or 6 months. Control performance significantly improved from 1 month to 6 months. Video-based inter-rater reliability revealed a good level of agreement for clinical OSCEs with a Cohen’s kappa of 0.632 (p<0.001) and a good level of agreement for surgical technical skills with a Cronbach’s alpha of 0.836.

The only significant differences in laparoscopic skills demonstrated between intervention and control groups were in accuracy using the 0 degree laparoscopic task at 6 months (90% versus 71.4, p = 0.030) and time using the 30 degree laparoscopic task at 6 months (129 seconds versus 177, p = 0.006).
**Table 8f: Primary Outcomes**

### CLINICAL OSCEs

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Post SIMweek</th>
<th>1 Month</th>
<th>6 Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
<td>Intervention</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>66 (23.3)</td>
<td>78 (9)</td>
<td>112 (5.75)</td>
<td>111 (7.25)</td>
</tr>
</tbody>
</table>

**Between group difference over time**

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>1 Month</th>
<th>6 Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>P = 0.181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td>P = 0.001</td>
<td>P = 0.007</td>
</tr>
</tbody>
</table>

**Within group difference over time**

<table>
<thead>
<tr>
<th>Post SIMweek v Baseline</th>
<th>1 Month v Post SIMweek</th>
<th>1 Month v Baseline</th>
<th>6 Month v 1 Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>P = 0.005</td>
<td>P = 0.646</td>
<td>P = 0.173</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td>P = 0.492</td>
<td>P = 0.043</td>
</tr>
</tbody>
</table>

### SURGICAL TECHNICAL SKILLS

<table>
<thead>
<tr>
<th>Total OSATS Score for 3 tasks</th>
<th>Baseline</th>
<th>Post SIMweek</th>
<th>1 Month</th>
<th>6 Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
<td>Intervention</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>48 (9)</td>
<td>52.5</td>
<td>78 (13.3)</td>
<td>78.5 (10.8)</td>
</tr>
</tbody>
</table>

**Between group difference over time**

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>1 Month</th>
<th>6 Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>P = 0.381</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Within group difference over time**

<table>
<thead>
<tr>
<th>Post SIMweek v Baseline</th>
<th>1 Month v Post SIMweek</th>
<th>1 Month v Baseline</th>
<th>6 Month v 1 Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>P = 0.005</td>
<td>P = 0.624</td>
<td>P = 0.092</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td>P = 0.011</td>
<td>P = 0.042</td>
</tr>
</tbody>
</table>
Figure 8e: (i) Clinical Performance - Clinical OSCEs

A box plot showing the clinical OSCE scores over different time periods (Baseline, 3 weeks, 1 month, 3 months, 6 months) for both intervention and control groups.
Figure 8e: (ii) Clinical Performance - Surgical Technical Skills
8.2.4.3: Secondary Outcome: (i) Confidence

Table 8g and Figure 8f display the results for self-reported confidence. Both intervention and control confidence significantly improved from baseline to 1 month and from 1 month to 6 months. However, the intervention group were significantly more confident than the control group at both 1 and 6-month assessments. There was no significant difference between post SIMweek and 1 month confidence for the intervention group.

8.2.4.4: Secondary Outcome (ii) Clinical Procedures Logbook

Table 8g displays the results for the clinical procedures logbook. No significant differences were observed between intervention and control total numbers of bedside procedures at any time point, although there was a non-significant trend towards a greater number in the intervention group at 4 months.
### Table 8b: Secondary Outcomes

#### SELF-REPORTED CONFIDENCE

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>101 (12.3)</td>
<td>95.5 (15)</td>
<td>149 (20.3)</td>
<td>129</td>
<td>142 (11)</td>
<td>148 (20.5)</td>
</tr>
<tr>
<td></td>
<td>(25.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Between group difference over time**

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention v Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.368</td>
<td></td>
<td>P &lt; 0.001</td>
<td></td>
<td>P = 0.022</td>
<td></td>
</tr>
</tbody>
</table>

**Within group difference over time**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Post SIMweek v Baseline</td>
<td>1 Month v Post SIMweek</td>
<td>1 Month v Baseline</td>
<td>6 Month v 1 Month</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.005</td>
<td></td>
<td>P = 0.005</td>
<td></td>
<td>P = 0.018</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td></td>
<td>P = 0.028</td>
<td></td>
<td>P = 0.028</td>
</tr>
</tbody>
</table>

#### CLINICAL PROCEDURES LOGBOOK

<table>
<thead>
<tr>
<th></th>
<th>1 Month</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
<td>Intervention</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>71 (52.8)</td>
<td>73.5</td>
<td></td>
<td>28.5 (16.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(29.5)</td>
<td></td>
<td></td>
<td>(8.25)</td>
<td></td>
</tr>
</tbody>
</table>

**Between group difference over time**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention v Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.958</td>
<td></td>
<td>P = 0.093</td>
<td></td>
<td>P = 0.713</td>
</tr>
</tbody>
</table>

*Last full week of work*
Figure 8f: Self-reported Confidence
8.2.5: Discussion

A one-week immersive simulation course significantly improved graduating medical students’ objectively measured clinical skills, surgical technical skills and self-reported confidence. This improvement persisted at 1 month and 6 month follow up assessments when compared to controls, despite no demonstrable differences in the volume of clinical experience. This suggests that improvements in training quality rather than increasing time in training can improve outcomes.

The SIMweek course accelerated subjects’ acquisition of clinical skills, technical skills and confidence so at the start of their Foundation Year 1 placement these were at a level similar to that seen at 1 month into Foundation Year 1. The intervention group’s clinical performance did not significantly improve from 1 month to 6 months, however their confidence continued to significantly improve and their technical skills demonstrated a trend towards improvement although this did not reach statistical significance. By contrast, the control group’s clinical skills did not display any significant improvement at 1 month compared to baseline, however their confidence significantly improved and their technical skills demonstrated a non-significant trend towards improvement. Control clinical skills, technical skills and confidence significantly improved from 1 month to the 6 month follow up period, while remaining lower than that of the intervention group. No significant differences were observed between intervention and control groups in numbers of clinical procedures performed at 1 month, 4 months and 6 months. The SIMweek course is thus a feasible and effective intervention to ease the transition from medical student to the first months as a clinician.
An interesting and perhaps surprising observation was the fact that intervention subjects’ performance did not significantly improve from post-course to 1 or 6 months, in contrast to the control group that did. This is likely due to the high performance of the intervention subjects on the post-course assessments. Due to their proximity to the top end of the assessment scale (median 112 out of a possible maximum of 119 for the clinical OSCE, 78 out of a maximum of 105 for the surgical technical skill and 149 out of a maximum 175 for self reported-confidence) there was in fact little room for improvement without a near-perfect performance. For future studies, perhaps more challenging tasks may allow the intervention subjects to demonstrate further improvement.

The results of this study are encouraging, however for an intervention to make a successful transition from the laboratory into clinical implementation, the costs of such a course must be considered. Table 8d describes the costs involved for a typical US training program to run a training scheme similar to the SIMweek course. As this inaugural course was run as a feasibility study it was run with intentionally small numbers to provide high faculty to student ratios. The costs per subject would be reduced further if run with larger numbers. The cost may prove a valuable investment when seen in the context of the cost of providing paid shadowing. In addition it is worth considering the potential effects this could have on the ‘August’ and ‘July’ effect and the subsequent cost and patient safety benefits. It must be remembered however that this study was not designed to evaluate effects on patient outcomes and future data on this would be of interest.
The controlled simulated environment allows subjects to learn from their mistakes without any compromise to patient care. While the clinical procedures used were the same as those used in standard medical school OSCEs, the immersive environment provides a realistic environment and allows for contextualised learning and practice of technical procedures. For example, a mock bleep scenario for poor urine output allows the subject to obtain a focused history and examination before coming to the conclusion that catheterisation is required and this can then be performed using a simulator. While shadowing an intern can also be useful, it relies on chance occurrences. The use of simulation can be one method to ensure coverage of relevant experience. Previous studies have shown promising results for the use of simulation to prepare medical students for the transition to surgical intern (Boehler et al. 2004; Brunt et al. 2008; Esterl et al. 2006; Okusanya et al. 2012; Peyre et al. 2006). These studies have often used self-reported confidence as their primary outcome (Esterl et al. 2006; Okusanya et al. 2012; Peyre et al. 2006). While confidence is likely an important factor in improving the anxieties of starting as a Foundation Doctor, this study adds significantly to the body of evidence by using objectively measured clinical performance as its primary outcome and providing data about skills retention 6 months following the intervention. Demonstration of persistently improved clinical performance may have a direct impact on patient care. This is not assessed in this study as all performance measurements were conducted in the simulated environment due to the described benefits. The potential effect of SIMweek on patient outcomes is an area for future work, where performance measurements should ideally be performed in the clinical environment.
The main limitation in the study design is the lack of randomisation. This was a pragmatic decision taken due to the inherent difficulties of recruiting to a study which requires commitment to 1 week of a subjects’ time during the holiday period between graduation and starting their clinical posts. In addition, due to the enthusiasm for the SIMweek course amongst the volunteers, it was anticipated that there would be a dropout bias in the control arm even if randomised. Previously reported ‘boot camp’ studies have also employed a non-randomised design (Okusanya et al. 2012). Where this study improves on this is by ensuring baseline assessments were completed for both control and intervention arms. Despite the non-randomised design; baseline scores between intervention and control arms revealed no significant differences in clinical skills, surgical technical skills or in confidence. Another limitation was the small number of controls. This was in part due to the need for baseline testing at a time when students are requested to curtail their holidays to start their compulsory shadowing. Since they were in the control arm there was not a clear benefit to participating although future simulation courses were offered to mitigate this. Two of the recruited controls had to be excluded as they were unable to commence their posts and one control was lost to follow up after 3 weeks. Despite the small numbers, the observed differences demonstrated statistical significance thus reducing the risk of a type II error. There is a risk that there was a difference between the two groups at baseline that was not statistically significant due to the small numbers. Nevertheless, the clinical and surgical technical skills performance was higher in the control group than the intervention group. Thus the improvement with the intervention may have been underestimated, therefore adding more weight to the rejection of the null
hypothesis. It must however be remembered that the secondary outcome of self-reported confidence was actually higher in the intervention group at baseline. The SIMweek curriculum used in this study was developed from expert consensus, facilitated by the literature review performed in the early part of this chapter. An essential next step to improve the curriculum would be to ensure it is refined using both evidence based methodology and importantly feedback from the end-users: the Foundation Doctors.

8.2.6: Conclusion

The immersive simulation course improved subjects’ objectively assessed clinical skills, surgical technical skills and self-reported confidence. Despite similar clinical experience as controls, the improved performance of the intervention group persisted at both 1 month and 6 months. This provides evidence that improvements in training quality rather than increasing time in training can improve leaning outcomes. This feasible and effective intervention to ease transition from student to junior doctor could reduce error and enhance patient safety.
Chapter 9: Conclusion
This concluding chapter starts with a reminder of the initial aims and hypotheses of this thesis.

9.1: Aims

- To identify variations in the quality of global surgical training programs

- To develop an assessment tool to objectively evaluate quality in surgical training programs.

- To develop and pilot interventions that will enhance the quality of surgical training programs.
9.2: Hypotheses

- There are variations in global surgical training programs.

- Quantity of training measured by time or caseload is not directly proportional to the quality of training.

- Quality of training can be measured.

- Training improvements can be made without increasing the quantity of training but by improving the quality of training.
9.3: Discussion

This thesis sought to investigate quality in surgical training from a global perspective. The surgical education of trainees has enjoyed much greater attention over the past decade. The emergence of new technology and the resultant advances in surgical techniques, combined with increased focus on patient safety outcomes have all contributed to raising the prioritisation of surgical training. However, while clinical outcomes, research outcomes and economic outcomes benefit from rigorous evaluation, surgical training outcomes are not routinely subjected to detailed analysis. Quantity of training in terms of number of hours worked, or number of operative cases performed, has formed the mainstay of training analysis. This thesis sought to shift the focus of training outcomes from quantity of training towards a focus on quality of training. With the increasing global trend towards mandatory work hour restrictions reducing the number of hours trainees can work, focusing on enhancing the quality of those training hours and making training more efficient is a crucial endeavour.

The introductory chapter of this thesis evaluated and summarised key developments in the history of surgical training within the UK. Postgraduate medical and surgical training has undergone significant structural reorganisation on a national level through numerous political interventions. The chapter summarised the structural changes to postgraduate surgical training in the UK. While surgical training is primarily an educational endeavour, unlike undergraduate university students, the surgical students are also in the frontline of providing a medical service. This chapter demonstrated the challenges faced when trying to balance the needs of the workforce and trainees when working within a system that provides a service to the majority
of a nation’s population. Postgraduate medical education has undergone a turbulent decade trying to reform its structure while incorporating European Working Time Directive principles and plan a workforce fit to deliver an efficient and high quality service once trained. Lessons have been learned from mistakes of the past and as a result postgraduate training has come under greater scrutiny.

In *Aspiring to Excellence* (Tooke 2008) the rationale and need for quality assurance processes in training was introduced. The *Temple* report (Temple 2010) also highlighted the need for established measures to evaluate the outcomes of training. In order to evaluate training quality, high quality training must be defined. Identifying variations in training is an important first step towards defining high quality training. Interest in surgical training has not been limited to the UK. Globally, there has been a drive to standardise the quality of surgical training through formalised national surgical training curricula. Chapter 2 compared the surgical training curricula of Canada, the UK, the USA, Australia and Hong Kong. Significant variations were demonstrated between the training programs in each country most notably in the length of their respective training programs. Postgraduate general surgical training takes twice as long in the UK with a minimum 10 years compared to 5 years in the USA and Canada. However, when the national surgical training curricula were objectively compared using with the General Medical Council’s ‘Standards for curricula and assessment’ (GMC 2010), few differences were demonstrated. The study reiterated the need for a robust method to assess the quality of surgical training afforded by different training programs.
In identifying some of the variations in global surgical training curricula, chapter 2 performed the first step towards defining high quality training. The two curricula that demonstrated some of the most marked variation were those of the UK and USA. These differences were further explored in chapter 3 by directly comparing training quality between 2 programs in the UK and the USA. While chapter 2 made comparisons by evaluating the structure and organisation of training programs using their national surgical training curricula, in chapter 3 comparisons were made through feedback from the end-user; the trainees of the respective training programs. An online questionnaire methodology was employed to seek direct feedback from the trainees regarding the nature and quality of their training. While the study findings would have been strengthened through a greater response rate and sampling more institutions, marked differences were again demonstrated between the respective training programs. The most marked difference was in the amount of supervision surgical trainees were receiving in the operating room, with trainees at the USA program heavily supervised with little experience of operating independently. The fact that trainees in the USA program received more teaching and better access to simulators are clearly strengths over the program in the UK. However the issue of supervision and its balance with independence is more difficult to attribute as a strength or a weakness.

The early studies in chapters 2 and 3 achieved the first aim of this thesis in identifying variability in surgical training programs. Furthermore they established that although there are clear differences in the quantity of hours worked and length of training programs, there may be more factors that contribute to the quality of training than time alone. Chapter 4 sought to
begin to address the second aim of the thesis, to develop an assessment tool to evaluate the quality of surgical training. The first step of this was to define the principles of high quality training. Qualitative methodology was employed to explore the subject of high quality training by conducting semi-structured interviews with general surgery trainees and trainers. The demographics of the clinicians interviewed were reflective of the UK surgical workforce and covered a wide range of experience of both trainees and trainers. This study provided data on the indicators of high quality surgical training and methods to measure them. For feasibility reasons, the interviews were conducted with UK clinicians with a view to using the information gleaned to develop international consensus in the subsequent study. The study participants were asked about interventions to improve the quality of training to inform the third aim of this thesis, to develop and pilot interventions to improve the quality of training. Trainee feedback about the training post was felt to be important which tied in with the main aim of the thesis to develop methods to assess the training quality of a program. Increasing exposure with the same trainer, particularly for supervised technical skills training, was identified as a priority for any intervention. Support was also voiced for the implementation of a structured teaching programme, which included practical skills in addition to the more traditional didactic teaching.

An interesting finding from chapter 4 was the demonstration of polarised views regarding whether all hospitals and consultants should train. The concept of potentially distinguishing faculty into educators who train from service clinicians, who do not, was very contentious. Some held strong views that all consultants should train the next generation of surgeons. A narrow majority of participants rejected this with equally strong views that the ability to train
was not a skill that everyone possessed or should be obliged to perform. Instead those with an interest should be better supported and permitted to do so by adjusting their workload accordingly. Furthermore, the study highlighted the need for faculty development and training with trainers explaining that they themselves would benefit from feedback and education about how to improve their teaching skills. A tool to objectively rate the quality of training could therefore highlight the skilled trainers and also provide constructive feedback to those trainers that are identified as needing to improve.

Chapter 5 sought to build on the results of the semi-structured interviews from chapter 4 and develop consensus from an internationally identified panel of surgical education experts on the indicators of high quality surgical training and how it can be measured. Seventy statements were developed based on themes arising from the results of chapter 4 and distributed to the panel of experts using an online questionnaire program. Fifty-three experts from eleven different countries participated in this modified Delphi study involving two rounds of voting. Pre-defined criteria for consensus and acceptance of statements resulted in 35 of the 70 statements being accepted. These were then used to develop the S-QAT Surgical training Quality Assessment Tool with descriptive anchors developed for each point in the quality scale.

The development of the S-QAT in chapter 5 was the first step towards robust evaluation of surgical training quality. It was developed through a rigorous process by employing established qualitative methodology techniques and involving a global community of surgical education experts. The natural next steps after development of a tool to assess training quality are to pilot
its use and to consider interventions to improve training quality based upon information gleaned from the development process. Chapter 6 describes further refinement and piloting of the S-QAT. The first part of chapter 6 describes an interactive workshop at an international surgical education conference in which the S-QAT was introduced and feedback was sought. The majority of delegates attending the workshop felt that surgical training quality should be measured. There was strong support for the results to be used to provide formative feedback while its use for summative feedback received only limited support. Constructive criticism of the S-QAT was taken on board, in particular regarding its length and the benefits of shortening it to improve completion.

The second part of chapter 6 builds on the findings of the workshop by piloting its use. While the workshop allowed subjects to observe the S-QAT tool and comment on its effectiveness, numbers were limited and subjects did not actually complete the S-QAT questions. The pilot of the S-QAT was implemented as part of a regional quality improvement initiative for surgical training in the North West London General Surgery training region. The S-QAT was administered as an online questionnaire to general surgery trainees and consultant trainers in the region. Responses were received from almost all training centers and almost half of all trainees and consultant trainers. The results supported the hypothesis that there is variability in surgical training quality, particularly in the provision of designated training operating lists. The balance between supervision and independence was identified as a strength of current training programs. Areas for improvement that were identified included the structured teaching programmes, training evaluation and the quality control of training.
The final chapters of this thesis address the third aim to develop and pilot interventions to improve training quality. One of the hypotheses of this thesis was that training improvements could be made by raising the quality of surgical training without necessarily increasing the quantity of training through hours worked. The introduction described The Temple report (Temple 2010) which recommended ‘making every moment count’ for training. This would support the notion that if more hours worked by trainees had training value then the quality could be improved without increasing the number of hours. The first intervention evaluated in this thesis to raise training quality is the concept of video based coaching. Moving on from the qualitative methodology in the earlier chapters, this intervention was evaluated by employing a randomised controlled trial methodology in chapter 7. Video based coaching addressed some of the themes from chapter 4. In defining high quality training, the semi-structured interviews identified that continuity between the trainee and the same consultant was an important indicator of training quality. This could have an added effect on the relationship between the trainer and trainee, which was another important indicator of high quality training. When discussing potential interventions to improve surgical training quality, there was support for increasing the continuity of exposure between a trainee and their consultant trainer. There was further support for supervised technical skills training. The video based coaching intervention could maximise the learning potential from every clinical case performed. In chapter 7 this is evaluated in a controlled simulation setting. Although subjects were medical students, they were first trained to proficiency with a competency based and validated virtual reality training curriculum before randomisation to the intervention or control
arms of the study. Despite equivalent exposure to practical laparoscopic skills training, the video based coaching intervention enhanced the quality of laparoscopic skills performance when compared to controls. This was first demonstrated in a virtual reality simulation setting and skills transfer to real tissue was demonstrated on cadaveric porcine laparoscopic cholecystectomies. The improvements did come at the expense of time but the improvement in quality was confirmed using three different validated rating scales.

While chapter 7 focused on improving the quality of predominantly technical skills training on an individual personalised level, the second intervention to improve training quality described in chapter 8 was aimed at institution or program level. When subjects were asked about potential interventions to improve the quality of surgical training during the semi-structured interviews in chapter 4, forty percent of subjects suggested the implementation of a structured teaching programme to include practical skills. In chapter 5 the presence of a structured teaching programme was identified as an important marker of training quality in the global Delphi study. Technical and non-technical skills teaching programmes were identified as areas of high variability in training quality in the pilot study of the S-QAT described in chapter 6. Therefore the intervention studied in chapter 8 was to design and pilot a structured teaching programme. The target stage of training for this intervention was the transition from medical student to Foundation doctor. The introductory chapter of this thesis highlighted this transition stage as an area of concern identified by a number of government reports. In *Aspiring to Excellence* (Tooke 2008) opinion had been divided on the subject of the adequacy of integration between the undergraduate curriculum and the newly implemented Foundation Curriculum.
The Tooke report recommended greater clinical experience towards the end of medical school training. In the recent *Shape of Training* report (Greenaway 2013) the transition from medical school to Foundation Year 1 was again highlighted as an area in need of attention and additional support.

The first section of chapter 8 develops a framework for an intensive week of simulation training to prepare graduating medical students for their first weeks as a Foundation Doctor. This type of structured teaching has been described as a ‘boot’ camp in the literature, and the systematic review of published literature conducted as part of the chapter provided an evidence base upon which to develop the framework. The resulting course is evaluated in a pilot study described in the second part of chapter 8. This was a non-randomised controlled trial by design owing largely to reasons of feasibility and subjects were followed up for 6 months of their Foundation Year training. Despite equivalent quantity of clinical experience in terms of time and no demonstrable differences in the quantity of clinical procedures performed, the intervention subjects’ objectively assessed clinical skills, surgical technical skills and self-reported confidence were significantly greater than the control subjects’ performance. Furthermore, this improved performance persisted at 1 month and 6 month follow up although the differences between the groups became narrower.

9.4: Limitations

The limitations of the studies in this thesis have been discussed in each chapter but the salient points are summarised again. In chapter 2, the UK’s General Medical Council standards were
used to compare the global surgical curricula. Since these are from the UK’s regulatory body, the UK surgical curriculum would naturally follow their guiding principles. In chapter 3, the limitations include those inherent to the methodology of employing a questionnaire study namely that the data collected was self reported from the trainees themselves and is therefore susceptible to recall bias. In addition the major limitation was that the trainee subjects were from a single training program in the UK and the USA and a greater response rate would have improved the validity of the results. Ideally the questionnaire study would have been expanded to a larger sample of training programs. While this would firstly require greater resources than were possible, expansion of the study would not necessarily have improved the response rate or eliminated recall bias.

The semi-structured interviews in chapter 4 were all conducted with UK surgeons. Furthermore, responses were immediate without allowing a period of time to consider the questions. However the benefit of this method was that it reduced the risk of subjects conferring with their colleagues and thus more accurately reflected their own personal views. Once again if resources permitted then semi-structured interviews with a global sample of surgeons would have enhanced the validity and generalisability of the findings. This limitation was addressed to some extent through the global panel of surgical education experts identified for the subsequent Delphi consensus study in chapter 5. Conversely, by employing a global panel of experts for the Delphi consensus study, there is a risk that the tool may in fact become too generalised and not sensitive enough to reflect local concerns of particular nations. The tool may indeed benefit from local development at a future stage to adapt to the local training
environment. In chapter 6 the S-QAT tool is piloted in the North West London training region. Since the tool was administered as an online questionnaire, it was again limited by its response rate although the response rate of 47% was better than in chapter 3. However, despite the improved response rate, the pilot study would be limited by a potential non-response bias.

The final chapters 7 and 8 describe intervention studies to try and improve the quality of surgical training. In chapter 7 the subjects are surgically naïve medical students. It would be preferable to use surgical trainees although in order to attempt to address this limitation, the subjects were trained to proficiency on a validated virtual reality laparoscopic cholecystectomy curriculum prior to randomisation. The main limitations to the study reported in chapter 8 are the lack of randomisation and the small number of controls. The non-randomised design was a pragmatic decision to aid recruitment while the control arm suffered from a greater loss to follow up.

In both studies described in chapters 7 and 8, the outcome assessments are performed in a simulated environment. In chapter 7 the porcine laparoscopic cholecystectomy assessments post intervention are performed in a simulated environment and ideally future studies with greater resources would evaluate the implementation of video based coaching in a clinical environment with surgical trainees. In chapter 8 the subjects are final year medical students and while they were followed for 6 months into their Foundation Year training, the assessments were still performed in a simulated environment. The assessment of training outcomes or performance is a large topic with much research being directed towards its improvement.
While assessing technical and non-technical performance both in the simulated and clinical settings is a pragmatic and sensible start to evaluating training outcomes, the optimum training outcomes will include patient related clinical outcomes and patient experiences.

9.5: Future Work

Further research should naturally seek to address the limitations of the work presented in this thesis. The preliminary work to define high quality surgical training has successfully led to the development of a tool to measure the quality of surgical training and to the development and pilot testing of a number of interventions to improve the quality of surgical training. Therefore future work should focus on both areas; to develop and implement the interventions, and to further refine the methodology to measure surgical training quality.

A major challenge for all surgical educational research is the translation of evidence-based innovations from the simulated setting into clinical practice. The video based coaching intervention reported in chapter 7 demonstrated encouraging results in the simulated setting but its implementation into clinical training requires careful consideration of the pressures of time. This was a driving factor in the decision to employ video based coaching using retrospectively reviewed videos of performance as oppose to real-time live coaching as it was felt this would enhance the flexibility of the process for both the trainees and trainers. Pragmatic modifications may be required to facilitate the clinical implementation of the video based coaching schedule using the GROW framework. This may involve weekly or monthly coaching sessions rather than immediately after the procedure and the potential effects of this
delay on the effectiveness of the coaching session would warrant further investigation. One method to reduce faculty time required for coaching is to encourage self-evaluation of the video performance prior to the coaching session. This could be formally evaluated by comparing self-evaluation using video feedback alone against video based coaching. A recent study reported that following a coaching intervention the accuracy of a trainee’s self evaluation was improved (Bonrath et al. 2015), therefore self-evaluation may well be a skill in itself that can be improved through coaching.

For the second of the interventions described in this thesis, SIMweek, the key next steps would be to focus on how the intervention can be refined and made more efficient in order to facilitate implementation to a larger scale that would be required if it were to be offered to all graduating medical students. Feedback from the students themselves was positive regarding the potential to reduce the intensity of the faculty:subject supervision ratio but this could benefit from more formal appraisal to ensure the benefits are not lost in a drive to lower the number of faculty training hours required. Furthermore the potential benefits in terms of patient safety and economic benefits of better-trained Foundation Doctors would warrant further analysis with a particular focus on patient outcomes. The concept of an intensive, immersive simulation course has proven to be effective for the transition from medical student to junior doctor. Further work should evaluate whether similar principles could be applied to other transitions in training, in particular the transition from core surgical trainee to the higher surgical trainee level.
Through a series of studies employing qualitative methodologies this thesis provides evidence that robust evaluation of surgical training quality is essential to raise overall standards of surgical training quality. Using a combination of trainers’ and trainees’ opinions and the consensus process using the international panel of surgical education experts, the assessment of training quality naturally adopted Donabedian’s 3-dimensional conceptual framework: structure, process and outcomes (Donabedian 1988). Surgical training quality can be assessed by examining the organisation in which it is delivered, thus examining the structure. The training techniques, including the trainers themselves and the quality of supervision and constructive feedback they provide, assess the process. The ultimate outcomes are the improvement in performance of the trainees and how this is reflected in their patients’ clinical outcomes. Future work should focus on developing these objective outcome measures to complement the structure and process measures assessed by the newly developed S-QAT. The final steps will be to develop appropriately weighted composite measures to assess surgical training quality. The Association of Surgeons in Training are currently working with an international Delphi Consensus Group seeking solutions to international challenges to surgical training. The robust evaluation of surgical training quality would be greatly facilitated by a global collaborative effort to support its continuing development.

Chapter 6 describes the first steps towards longitudinal assessment of surgical training quality. Continuous and dynamic refinement of the S-QAT tool will be necessary to improve the assessment of surgical training quality. Further and larger pilot studies will facilitate the refinement of the S-QAT and permit long-term analysis of the effects of measuring surgical
training quality. The objective measurement and maintenance of training quality standards is relatively novel in the context of surgery. However, quality standards have been well established in the education sector for many years now. In England, the Office for Standards in Education (OFSTED) is responsible for independent external evaluation of a school’s effectiveness. The overall objective is to identify strengths and weaknesses of a school and promote improvement where necessary by providing clear recommendations. Similar to the S-QAT, OFSTED inspections analyse feedback from both the trainers (teaching faculty) and trainees (school pupils). Evaluations of schools are performed on a longitudinal basis with the frequency of inspections increased where poor performance has been highlighted. Pupil performance and in particular their performance in comparison with national averages forms part of the assessment. The equivalent when measuring surgical training quality in a training institution or deanery would be challenging, as surgical trainee performance is difficult to define. Exam performance in the FRCS examination would provide an indication of knowledge and to some extent clinical skills, however only limited technical skills assessment are currently performed and in a non-standardised manner. Nevertheless the framework for the assessment of the quality of education has been established in the schools and university sector. Future surgical training quality evaluation would benefit from learning from the experience of government departments such as OFSTED.
9.2: Concluding Remarks

The variability in global surgical training programs has been described. Through a series of qualitative methodologies, high quality surgical training has been described. The products of these studies include the development of an assessment tool to evaluate the quality of training within a surgical training institution. This tool has subsequently been piloted in the North West London training region where variability has been confirmed. Defining high quality surgical training has facilitated the development of ideas for interventions to improve the quality of surgical training. Two interventions have been developed and piloted with encouraging results.

Quality in surgery has enjoyed much greater exposure in recent years. Processes have been developed to thoroughly evaluate clinical outcomes and research output. These processes are dynamic with regular refinement to improve the validity of the assessment. Quality in surgical training has largely proceeded without structured assessment. Where quality of training has been examined this is often performed through operative case volume or duration of training. These are surrogate markers for quality, with an assumption that increased volume of cases or hours worked leads to high quality training.

Some key issues that need to be addressed to enhance the quality of surgical training include:

- Structured and robust assessment of surgical training quality is essential.
  - This thesis demonstrates almost unanimous agreement among the profession that the evaluation of surgical training quality will have a positive impact and raise standards.
As with all quality assessment programmes, the methodology developed in this thesis will require continuous refinement in order to continue to improve the robustness of the evaluations.

- Structured teaching programmes for technical, non-technical and research skills need to be improved.
  - Much of surgical training takes place in the workplace but there is a need to develop structured teaching programmes to augment this learning.
  - Simulation facilities can be valuable in these teaching programmes but the possession of equipment alone will not make the programmes successful.
- Conflict between service provision and surgical training needs to be addressed.
  - The two cannot be completely separated since there is no end-point to surgical training; it is a lifelong endeavour for a surgeon to maintain and improve their professional skills.
  - Nevertheless the question of whether all hospitals and all surgeons should train has been highlighted as a contentious one.
  - The Temple report (Temple 2010) noted that consultants that accept greater responsibility for training are often not appropriately recognised or rewarded for their efforts.
  - If service provision clinicians separate themselves from faculty educators then innovative ‘productivity’ measures will need to include more than the traditional measures of service provision alone.
- Faculty educators require adequate resources to train the next generation of surgeons.
This must include faculty training and feedback on their teaching ability.

Chapters 4 and 5 confirmed that trainers should be provided feedback on their training ability.

The NHS Future Forum (DoH 2012) recommended that clinicians delivering training should be appropriately supported both in terms of their development as a trainer and in providing enough time to train.

Institutions must recognise that faculty educators need extra time or a reduced clinical workload to allow trainees to be taught in a safe environment.

- Healthcare outcomes will be improved through higher quality surgical education.

- Healthcare has remained a huge political issue with intense scrutiny of healthcare outcomes.

- This is exemplified at the:
  - Institutional level with the Francis report into the concerns at the Mid Staffordshire NHS Foundation Trust (Francis 2013).
  - Individual level with an increasing pressure for publication of individual surgeon outcomes data in the UK.

- Training can drop down the list of political priorities as the clinical benefits are not necessarily seen immediately.

- However, the Tooke inquiry (Tooke 2008) noted that there was evidence of a positive correlation between Trusts that invest in education and research and Healthcare Commission ratings. The Temple report (Temple 2010) also noted that highly regarded training institutions also delivered high quality clinical care.
It is crucial to emphasise the importance of high quality training to maintain current standards of care and also to improve clinical care in the future.

It is therefore essential to have protected funding for surgical education.

- Analysis of healthcare outcomes in terms of patient safety and patient experiences will reveal areas in need of improvement. This information can be used to focus training improvements.

- A recent successful example of this is the introduction of laparoscopic colorectal surgery. In 2006 NICE recommended laparoscopic resection should be offered as an alternative to open surgery for colorectal cancer. There were not enough trained laparoscopic surgeons to allow the recommendations to be implemented. Laparoscopic colorectal activity was less than 5% at the time and as a result the LAPCO national training programme was setup to train enough colorectal surgeons to offer a laparoscopic service (Coleman).

This thesis lays the foundations for robust evaluation of surgical training quality. Variability in training quality has been demonstrated. The interventions described and piloted illustrate the potential to improve training through evidence based improvements in training quality rather than relying on simply increasing time in training with the hope of increased exposure to operative case load. Evidence based medicine has become the gold standard for patient centred care, it is now time to practice evidence based surgical training to ensure surgeons of the future are trained to the highest standards and in the most efficient manner.
References
ABS 'The American Board of Surgery Training Requirements for General Surgery Certification',

ACGME (2010), 'RRC News Surgery Resident's Update Fall 2010'.
--- (2011a), 'Resident Duty Hours in the Learning and Working Environment: Comparison of 2003 and 2011 Standards'.
--- (2011b), 'The ACGME 2011 Duty Hour Standards: Enhancing Quality of Care, Supervision and Resident Professional Development'.
--- (2012), 'ACGME Program Requirements for Graduate Medical Education in General Surgery'.

ACGME. and ABS. (2014), 'The General Surgery Milestone Project - A joint Initiative of The Accreditation Council for Graduate Medical Education and The American Board of Surgery'.

ACS 'Bulletin of the American College of Surgeons: Looking forward - February 2013',

Aggarwal, R., et al. (2004), 'The simulated operating theatre: comprehensive training for surgical teams',
*Qual Saf Health Care*, 13 Suppl 1, i27-32.


Anthony, A., Cox, M., and Cullingford, G. 'General Surgery SET 1 competencies',


BBC 'Formula 1 simulation affects real world performance',


BMA 'British Medical Association: European Working Time Directive for Junior Doctors',


Coleman, M. 'LAPCO National Training Programme in Laparoscopic Colorectal Surgery',


CSHK 'The College of Surgeons of Hong Kong: web site',


--- (2004), 'The College of Surgeons of Hong Kong: An outline of the required competencies of a graduating trainee in General Surgery'.

--- (2013), 'The College of Surgeons of Hong Kong: New curriculum for basic surgical training'.

CUG 'The Complete University Guide: University League Table 2012',


DFRL (2011), 'Dr Foster Research Limited: Inside your hospital. Dr Foster hospital guide 2001-2011'.

--- (2012), 'Dr Foster Research Limited: Fit For The Future? Dr Foster hospital guide 2012'.

DoH 'Department of Health: Shadowing for Foundation Year One Doctors Introduced',

--- (2003), 'Department of Health: The response of the four UK Health Ministers to the consultation on Unfinished Business: Proposals for reform of the Senior House Officer grade'.
--- (2004), 'Department of Health: Modernising Medical Careers - The next steps'.
--- (2008a), 'Department of Health: A High Quality Workforce NHS Next Stage Review'.
--- (2008b), 'Department of Health: High Quality Care For All - NHS Next Stage Review Final Report'.
--- (2012), 'Department of Health: Education and Training - next stage. A report from the NHS Future Forum'.


Donaldson, L (2002), 'Unfinished business: proposals for reform of the senior house officer grade - a paper for consultation'.

Donaldson, L. (2009), '150 Years of the Annual Report of the Chief Medical Officer', in Department of Health (ed.).


Dzau, V. J., et al. (2010), 'The role of academic health science systems in the transformation of medicine', Lancet, 375 (9718), 949-53.


FC 'The Fellowship Council - Universal Fellowship Application and Match Service',


Fletcher, K. E., et al. (2005), 'Effects of work hour reduction on residents' lives: a systematic review', *JAMA*, 294 (9), 1088-100.


FP (2011), 'The Foundation Programme Core Procedures & Assessment Forms'.

--- (2012), 'The Foundation Programme Curriculum 2012'.


GMC (1997), 'General Medical Council: The New Doctor Recommendations on General Clinical Training'.

--- (2010), 'General Medical Council: Standards for curricula and assessment systems'.


--- (2011), 'General Surgeons Australia: Training regulations handbook - For the surgical education and training program in General Surgery'.

310


HEFCE (2007), 'Higher Education Funding Council for England: 116th Meeting of the Board, 20 September 2007: Minutes of previous meeting (M507)', Page 5 in Minutes of previous meeting (M507)


--- (2010), 'The Intercollegiate Surgical Curriculum: Educating the surgeons of the future. General Surgery syllabus '.


JCST (2012a), 'Joint Committee on Surgical Training: Quality Indicators and Trainee Survey'.

--- (2012b), 'Joint Committee on Surgical Training: Guidelines for the award of a CCT in General Surgery'.

Jen, M. H., et al. (2009), 'Early In-Hospital Mortality following Trainee Doctors' First Day at Work', Plos One, 4 (9).


Krajewski, A., et al. (2013), 'Implementation of an intern boot camp curriculum to address clinical competencies under the new Accreditation Council for Graduate Medical Education supervision requirements and duty hour restrictions', JAMA Surg, 148 (8), 727-32.


NASA (2014), 'Coaching and Mentoring',


Nicholl, D. J. (2007), 'All the president's men and modernising medical careers', *Lancet*, 369 (9574), 1687.


RCPSC 'The Royal College of Physicians and Surgeons of Canada: web site',


--- (2010a), 'The Royal College of Physicians and Surgeons of Canada: Objectives of training in the specialty of General Surgery'.

--- (2010b), 'The Royal College of Physicians and Surgeons of Canada: Specialty training requirements in general surgery'.

--- (2010c), 'The Royal College of Physicians and Surgeons of Canada: Specialty training requirements in surgical foundations'.

--- (2010d), 'The Royal College of Physicians and Surgeons of Canada: Objectives of surgical foundations training'.

--- (2010f), 'The Royal College of Physicians and Surgeons of Canada: Specific standards of accreditation for residency programs in General Surgery'.

--- (2011), 'The Royal College of Physicians and Surgeons of Canada: General standards of accreditation'.


REF 'Research Excellence Framework: Panel criteria and working methods',


SCORE 'Surgical Council on Resident Education: General Surgery Residency Curriculum Portal',


--- (2011), 'Surgical Council on Resident Education: General Surgery Residency Curriculum Outline'.


Tooke, J. (2008), 'Aspiring to Excellence - Findings and Final Recommendations of the Independent Inquiry into Modernising Medical Careers', in Department of Health (ed.).

UKCRC. and MMC 'UK Clinical Research Collaboration & Modernising Medical Careers: Medically and dentally qualified academic staff - Recommendations for training the researchers and educators of the future'.


Volpp, K. G., et al. (2012), 'Effect of a protected sleep period on hours slept during extended overnight in-hospital duty hours among medical interns: a randomized trial', JAMA, 308 (21), 2208-17.

Whitmore, J. (1992), Coaching for Performance (Nicholas Brealey Publishing).


Appendices
There are global variations in surgical training programs. The most marked of these are the differences in the length of the training programs. In the USA post-graduate surgical training is completed in a 5-year surgical residency, while in the UK it takes a minimum of 10 years post graduate training to reach Consultant/Attending level. Despite these variations, the common end product of these training programs is a trained surgeon capable of independent practice. It follows that if variations exist and yet the end product is similar then there are potentially variations that can improve the quality and efficiency of surgical training.

The current move away from volume based training to competency-based training has led to increased interest in the quality of surgical training. In the UK, the Joint Committee on Surgical Training (JCST) has developed quality indicators for surgical training. In the USA, similar indicators can be inferred from the American Board of Surgery (ABS) and Accreditation Council for Graduate Medical Education (ACGME) Surgical Residency program requirements. The purpose of this survey is to compare surgical training between the UK and USA based on questions developed from the themes of these quality indicators.

Study Investigators:

IMPERIAL COLLEGE LONDON
Dr. Pritam Singh, MBBS MA MRCS
Professor Ara Darzi, MD KBE FmedSci FRCS FACS
Dr. Rajesh Aggarwal, MBBS MA PhD FRCS

UNIVERSITY OF PENNSYLVANIA
Mr. Daniel Hashimoto, BA
Dr. Noel Williams, MBCh FRCSi FRCS

*INFORMED CONSENT*

This study survey conveys minimal risk to you. No confidential or personal identifying information is collected. The survey should take you no longer than 5-10 minutes to complete, thus minimizing risk of fatigue to you. Participation in this study conveys no direct benefit to you. By taking this survey, you are consenting to have the data from your anonymous responses reviewed by study investigators.

Thank you in advance for your participation.

Do you agree to participate in this survey?

☐ Yes ☐ No

*Operative Experience and Workload*

*Please answer the following demographic questions*

<table>
<thead>
<tr>
<th>Country</th>
<th>Gender</th>
<th>Age</th>
<th>Year of Post-Graduate Training</th>
<th>Current status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Surgical Education Survey

**In a typical week, how many cases do you participate in as the...**

<table>
<thead>
<tr>
<th></th>
<th>Elective cases</th>
<th>Emergency cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary surgeon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First assistant</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**When operating as the primary surgeon what level of supervision do you receive in ELECTIVE CASES?**

*Answers must add up to 100%*

- Attending/consultant off-site but available
- Attending/consultant on-site but not in operating room
- Attending/consultant in the operating room but not scrubbed
- Attending/consultant scrubbed

**When operating as the primary surgeon what level of supervision do you receive in EMERGENCY CASES?**

*Answers must add up to 100%*

- Attending/consultant off-site but available
- Attending/consultant on-site but not in operating room
- Attending/consultant in the operating room but not scrubbed
- Attending/consultant scrubbed

**When operating, how often do you take an active role in the pre-operative briefing (i.e. the "time-out")?**

- [ ]

**What proportion of the time do you spend during a typical on call DURING THE DAY...**

*Answers must total 100%*

- On a Consultant/Attending led ward round
- Reviewing new admissions/referrals/consultations
- Reviewing surgical in-patients
- In the operating theatre
- Elective commitments (e.g. clinic)
- Resting
### Surgical Education Survey

**What proportion of the time do you spend during a typical on call AT NIGHT...**

[Answers must total 100%]

- On a Consultant/Attending led ward round
- Reviewing new admissions/referrals/consultations
- Reviewing surgical in-patients
- In the operating theatre
- Elective commitments (e.g. clinic)
- Resting

**In a typical week, how many sessions (1 session = 1 half-day) do you spend:**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>In theatre</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In clinic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**In clinic, what proportion of new versus follow-up patients do you see?**

- [ ] Highly New
- [ ] Mostly New
- [ ] Mixed
- [ ] Mostly Follow-up
- [ ] Highly Follow-up

**How often do you have the opportunity to participate in a gastrointestinal endoscopy session?**

- [ ] Weekly
- [ ] Every second week
- [ ] Every third week
- [ ] Every fourth week
- [ ] Rarely

**Do you play an active role in clinical governance (i.e. improving the quality of patient care within your health system)?**

- [ ] Yes
- [ ] No

**Education**

**How many hours of scheduled teaching time do you have per week?**

- [ ] 1
- [ ] 2
- [ ] 3
- [ ] 4
- [ ] 5
- [ ] 6
- [ ] 7
- [ ] 8
- [ ] 9
- [ ] 10

**How many of those teaching hours are protected (i.e. free from clinical responsibility)?**

- [ ] 1
- [ ] 2
- [ ] 3
- [ ] 4
- [ ] 5
- [ ] 6
- [ ] 7
- [ ] 8
- [ ] 9
- [ ] 10

**What types of training tools do you have access to? [Check all that apply]**

- [ ] Virtual reality laparoscopy trainers (e.g. LapMentor, LapSim)
- [ ] Endoscopy trainers
- [ ] Laparoscopy box trainers (e.g. FLS box)
- [ ] Human patient simulators (e.g. SimMan)
- [ ] Cadaveric animals
- [ ] Live anaesthetised animals
- [ ] Human cadavers
- [ ] None
### Surgical Education Survey

**Does your program enable access to any of the following:**

| Study leave (e.g. to attend conferences) |  |
| Management training course (e.g. leadership or business training) |  |
| Teaching and Education courses (e.g. "Train the trainers") |  |
| Opportunities to update ATLS certification |  |

**Is lab-based (e.g. skills centre) training compulsory in your current training year?**

- [ ] Yes
- [ ] No

*You answered that lab-based (skills center) training is compulsory in your training year.*

**How often do you attend lab-based training?**

(please fill in the appropriate box)

- If weekly, hours/week
- If monthly, days/month
- If annually, weeks/year

*You answered that lab-based (skills center) training is not compulsory in your training year.*

**How often do you spend practicing operative skills on simulators?**

(please fill in the appropriate text box)

- Hours/week
- Days/month
- Weeks/year

**Is your non-compulsory lab-based training...**

- [ ] Self-funded
- [ ] Industry-funded
- [ ] Department-funded
- [ ] Not Applicable
## Surgical Education Survey

### Evaluation and Assessment

**Do you have a formalized objective setting session (i.e. setting educational goals) with a consultant/attending at the start of each placement/rotation?**

<table>
<thead>
<tr>
<th>Session</th>
<th>Type</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**How often do you typically receive feedback on your technical performance...**

[Please rate from 1 (never) to 5 (always).]

<table>
<thead>
<tr>
<th>Session</th>
<th>Verbal feedback</th>
<th>Written feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>During an operation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediately after an operation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>End of placement/rotation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**In a typical year, how many written feedback evaluations (workplace-based assessments if in the UK) would you expect to receive?**

<table>
<thead>
<tr>
<th><strong>Research</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you taken time off during your training to conduct research?</td>
</tr>
<tr>
<td>[Please answer yes if currently in research]</td>
</tr>
<tr>
<td>○ Yes</td>
</tr>
</tbody>
</table>

### Research time not yet taken

**How much time off do you intend to take to conduct research?**

<table>
<thead>
<tr>
<th><strong>Research</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is it compulsory?</td>
</tr>
<tr>
<td>○ Yes</td>
</tr>
</tbody>
</table>

**Approximately what percentage of your time will be devoted to clinical duties (i.e. operating theatre, ward rounds, locum/"moon lighting") while you conduct your research?**

<table>
<thead>
<tr>
<th><strong>Research</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is it compulsory?</td>
</tr>
<tr>
<td>○ Yes</td>
</tr>
</tbody>
</table>

### Evaluation and Assessment

<table>
<thead>
<tr>
<th>Session</th>
<th>Type</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Research</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you taken time off during your training to conduct research?</td>
</tr>
<tr>
<td>[Please answer yes if currently in research]</td>
</tr>
<tr>
<td>○ Yes</td>
</tr>
</tbody>
</table>

### Research time not yet taken

**How much time off do you intend to take to conduct research?**

<table>
<thead>
<tr>
<th><strong>Research</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is it compulsory?</td>
</tr>
<tr>
<td>○ Yes</td>
</tr>
</tbody>
</table>

**Approximately what percentage of your time will be devoted to clinical duties (i.e. operating theatre, ward rounds, locum/"moon lighting") while you conduct your research?**
**Surgical Education Survey**

*After what year of training do you plan to begin your research?*

*What do you expect will be the outcome of your research?*

[Check all that apply]

- Master's Degree (MS, MSc, MBA, etc.)
- PhD
- Juris doctorate (JD)
- Medical Doctorate [UK]
- Published journal articles
- Presentations at scientific meetings
- Additional income

Other (please specify)

*Will your research be funded and what do you expect will be the source of your funding?*

<table>
<thead>
<tr>
<th>Funding guaranteed?</th>
<th>Source of funding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Research time taken**

*How much time off have you taken to conduct research?*

*Was it compulsory?*

- Yes
- No

*Approximately what percentage of your time was devoted to clinical duties (i.e. operating theatre, ward rounds, locum/"moon lighting") while you conducted your research?*

*After what year of training did you begin your research?*
Surgical Education Survey

*What was the outcome of your research?*
[Check all that apply]

- Master’s Degree (MS, MSc, MBA, etc.)
- PhD
- Juris doctorate (JD)
- Medical Doctorate [UK]
- Published journal articles
- Presentations at scientific meetings
- Additional income
- Other (please specify)

*Was your research funding guaranteed, and what was the source of your funding?*

<table>
<thead>
<tr>
<th>Funding guaranteed?</th>
<th>Source of funding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

End

Thank you very much for your participation in this survey. If you have any questions, please feel free to contact the study investigators.

Dr. Pritam Singh  
pritam.singh@imperial.ac.uk  
Imperial College London

Daniel Hashimoto  
dhash@mail.med.upenn.edu  
Perelman School of Medicine  
University of Pennsylvania
### Appendix B: Quality of Surgical Training Semi Structured Interview Protocol

<table>
<thead>
<tr>
<th>Topic</th>
<th>Key Points/Phrases</th>
</tr>
</thead>
</table>
| 1. Introduction. | • Confirm identity of subject and their details  
  o Trainer/trainee  
  o Years as a consultant/Training level  
  • Confirm permission to record the interview  
  • Interviewer introduction and purpose of project |

“Recent changes to educational commissioning will mean that trusts have to demonstrate high quality of surgical training in order to be allocated trainees i.e. the funding will follow the trainee. The question remains how can a trust demonstrate high quality of surgical training? Our goals are firstly to identify  
  o What constitutes high quality of surgical training  
  o Secondly to define how we can measure this  
  o and finally to think about how we can improve the quality of surgical training”

<table>
<thead>
<tr>
<th>2. Define the problem.</th>
<th>Opening statement:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“I would like to start by asking about surgical training in general.”</td>
</tr>
<tr>
<td></td>
<td>“In your opinion do you think there is variability in the quality of surgical training or to put it another way: are some jobs better training posts than others?”</td>
</tr>
<tr>
<td></td>
<td>“Could you please elaborate on that? Is it simply a case of the quantity of cases or hours worked or can you think of other factors?”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Identify learning moments across the surgical training pathway. Vertical/horizontal (longitudinal)</th>
<th>“Ok now that we have talked about the issue in general, I would like to spend some time focusing on some of your learning and training experiences”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“In your opinion where are the potential opportunities for learning in the workplace or ‘on the job training opportunities’?”</td>
</tr>
</tbody>
</table>
|                                                                                               |   o Theatre  
|                                                                                               |   o Ward round  
|                                                                                               |   o Clinic |
|                                                                                               | “What opportunities are there for learning outside the workplace?” |
|                                                                                               |   o Lectures/Conferences  
|                                                                                               |   o Workshops  
|                                                                                               |   o Courses |
4. Identify indicators of high quality surgical training.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Simulation</strong></td>
</tr>
</tbody>
</table>
| “We have talked in general about the difference between a good job and a bad job? Can you tell me about your best and worst training jobs/trainees?**  
   • They do not have to have been your own experiences but perhaps those you have heard about from colleagues” |
| “What factors in particular made the job/trainee so good or bad?” |
| “If you lived in an ideal world free of the limitations of the current system. Please describe what you would look for in an ideal training job? i.e. describe your wish list” |
| “So now suppose you were trying to choose between a number of training jobs for yourself [or one of your good trainees]. What indicators of quality would you want more information on to help you decide which one is the best training job?”  
   • “These indicators may not be available but answer what you would ideally want to know.” |
| Examples/Prompts:  
   • Supervision  
     o Should everything be supervised?  
       * we can always learn more  
     o Or is there a benefit in reducing supervision?  
       * learn from your struggles  
   • Feedback  
     o Written or verbal and how often?  
     o From who? * Trainer only  
       * Or multi-source/Nursing Staff/Peers  
     o One consultant v multiple consultants  
   • Organisational  
     o Shift patterns  
     o Blocks of emergency/elective training  
     o Trainer/Trainee ratios  
     o Presence/absence of nurse practitioners/surgical assistants  
     o Annual/Study leave/Zero week  
       • Is more better?  
   • Simulation  
     o Accessibility  
     o Costs  
     o Frequency: Intensive/Distributed  
     o Compulsory or voluntary  
   • Research/Audit  
   • Non-technical skills/Professional Development |
### 5. How do we measure high quality training?

“We have talked about the ideal high quality training job but how feasible do you think it is to measure the quality of surgical training?”

*Draw on the indicators they mention in question 4*

“Earlier you described some markers of a high quality surgical training program, how can these factors be measured objectively?”

- Logbooks
  - Supervision
  - Feedback/Assessments
- Examinations and associated pass rates/Dropout rates
- Destination of graduates from the training program
- External audit of organisational features

“Who should perform the assessments?”

- Trainee feedback - *consumers*
- Trainer feedback - *providers*
- Independent/Peer review
- Paper based/Web based/An ‘App’ for real time feedback

“Should the results be made freely available”

- League tables?

“What do you think the purpose of measuring the quality of surgical training is or to put it another way: will measuring the quality of surgical training have any positive or negative effects?”

*Competition will improve standards/will it turn into another tick-box exercise?*

### 6. Identify potential ways to improve the quality of surgical training

“We have talked about some of the factors that would make a high quality surgical training program. Some of these factors are likely to be uncontrollable. However, are there any things we could realistically do to improve the quality of surgical training?”

*Draw on their previous answers*

“So for each of the indicators of quality you have mentioned, can you think of at least one way to improve it?”

### 7. Who should carry out surgical training?

“Do you think all Hospitals and Consultants should have a duty to train their juniors”

- Centralise training?
- or accreditation of trainers within departments?

“Is there a role for dedicated trainers with no clinical commitments?”

“And finally we have discussed surgical training quality and how to improve it but whose responsibility do you think it is to ensure the quality of surgical training?”
| 8. Further questions or comments | “Do you have anything further you wish to add?” |
## Appendix C: Statements not reaching pre-determined criteria for acceptance

<table>
<thead>
<tr>
<th>STEM</th>
<th>STATEMENT</th>
<th>MEAN</th>
<th>SD</th>
<th>MISING DATA %</th>
<th>% RATING 4 or 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Defining High Quality Surgical Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision versus Independence</td>
<td>All operative cases should be supervised with the trainer present inside the operating room.</td>
<td>2.60</td>
<td>1.22</td>
<td>0.00</td>
<td>25.71</td>
</tr>
<tr>
<td>Good feedback from previous trainees is an important marker of surgical training quality</td>
<td></td>
<td>4.00</td>
<td>0.69</td>
<td>0.00</td>
<td>77.14</td>
</tr>
<tr>
<td>Continuity</td>
<td>The degree of continuity between a trainee and a trainer is an important marker of surgical training quality.</td>
<td>3.46</td>
<td>0.78</td>
<td>0.00</td>
<td>40.00</td>
</tr>
<tr>
<td></td>
<td>Training rotations should be a minimum of 6 months in duration.</td>
<td>3.40</td>
<td>1.19</td>
<td>0.00</td>
<td>51.43</td>
</tr>
<tr>
<td>The balance between service commitments and training is an important marker of surgical training quality</td>
<td></td>
<td>3.86</td>
<td>0.73</td>
<td>0.00</td>
<td>71.43</td>
</tr>
<tr>
<td>Constructive Feedback</td>
<td>Regular (daily) informal feedback is essential.</td>
<td>3.38</td>
<td>0.84</td>
<td>0.03</td>
<td>44.12</td>
</tr>
<tr>
<td></td>
<td>Regular (monthly) formal feedback is essential.</td>
<td>3.83</td>
<td>0.79</td>
<td>0.00</td>
<td>71.43</td>
</tr>
<tr>
<td>Structure and Organisation</td>
<td>Structure &amp; organisation of the timetable can determine surgical training quality.</td>
<td>3.86</td>
<td>0.81</td>
<td>0.00</td>
<td>71.43</td>
</tr>
<tr>
<td></td>
<td>Trainees should be on call with their own firm/team.</td>
<td>3.43</td>
<td>0.85</td>
<td>0.00</td>
<td>54.29</td>
</tr>
<tr>
<td></td>
<td>The timetable should make it clear which team members are present well in advance of operating lists or other training opportunities.</td>
<td>3.83</td>
<td>0.79</td>
<td>0.00</td>
<td>77.14</td>
</tr>
<tr>
<td>Personalised training</td>
<td>It is essential for training courses to be tailored to the individual trainee’s level.</td>
<td>3.69</td>
<td>0.87</td>
<td>0.00</td>
<td>60.00</td>
</tr>
<tr>
<td>Simulation and Teaching</td>
<td>Access to surgical simulation equipment is a marker of surgical training quality.</td>
<td>3.86</td>
<td>0.81</td>
<td>0.00</td>
<td>65.71</td>
</tr>
<tr>
<td></td>
<td>Presence of supervised training on surgical simulators is a marker of surgical training quality.</td>
<td>3.63</td>
<td>0.81</td>
<td>0.00</td>
<td>60.00</td>
</tr>
<tr>
<td>Teamwork</td>
<td>The competition for training in terms of the trainer to trainee ratio affects the quality of surgical training.</td>
<td>3.62</td>
<td>0.87</td>
<td>0.03</td>
<td>64.71</td>
</tr>
<tr>
<td></td>
<td>Seniority of surgical colleagues affects training quality.</td>
<td>3.51</td>
<td>0.78</td>
<td>0.00</td>
<td>51.43</td>
</tr>
<tr>
<td></td>
<td>Seniority of multi-disciplinary team members affects the quality of surgical training.</td>
<td>3.46</td>
<td>0.78</td>
<td>0.00</td>
<td>45.71</td>
</tr>
<tr>
<td>The opportunity to attend training courses is a marker of surgical training quality</td>
<td></td>
<td>3.80</td>
<td>0.63</td>
<td>0.00</td>
<td>68.57</td>
</tr>
<tr>
<td>The balance of emergency workload versus elective workload is a marker of surgical training quality</td>
<td></td>
<td>3.49</td>
<td>0.92</td>
<td>0.00</td>
<td>48.57</td>
</tr>
</tbody>
</table>
## Measuring the Quality of Surgical Training

<table>
<thead>
<tr>
<th>Operating Room</th>
<th>3.34</th>
<th>1.08</th>
<th>0.00</th>
<th>54.29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainees should be involved in selection of cases that are placed on a theatre list.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Logbook analysis

<table>
<thead>
<tr>
<th></th>
<th>3.77</th>
<th>0.84</th>
<th>0.00</th>
<th>68.57</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainee logbook analysis is essential to measure surgical training quality.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregate measures of trainee logbooks should be used.</td>
<td>3.51</td>
<td>0.70</td>
<td>0.00</td>
<td>51.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3.66</th>
<th>0.64</th>
<th>0.00</th>
<th>57.14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainer logbook analysis is essential to measure surgical training quality: case load versus how much their trainees do.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logbooks should include clinic numbers in addition to operative cases.</td>
<td>3.54</td>
<td>0.82</td>
<td>0.00</td>
<td>57.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3.77</th>
<th>0.81</th>
<th>0.00</th>
<th>65.71</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logbooks should be modified to allow greater detail about how much of the operation a trainee has performed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### The effect of timetable/rota structure on:

<table>
<thead>
<tr>
<th></th>
<th>3.71</th>
<th>0.71</th>
<th>0.00</th>
<th>68.57</th>
</tr>
</thead>
<tbody>
<tr>
<td>The continued interaction between the trainee and their trainer should be assessed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The continuity of care for patients should be assessed.</td>
<td>3.83</td>
<td>0.79</td>
<td>0.00</td>
<td>71.43</td>
</tr>
</tbody>
</table>

### Multi-source feedback to measure surgical training quality?

<table>
<thead>
<tr>
<th></th>
<th>3.37</th>
<th>0.91</th>
<th>0.00</th>
<th>51.43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent (third party) feedback is an essential component.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Data on training quality should be sent to a central administration unit to collate and process the results

<table>
<thead>
<tr>
<th></th>
<th>3.47</th>
<th>0.95</th>
<th>0.03</th>
<th>50.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective numbers such as case load can be provided by hospital administrators or national databases such as Hospital Episode Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Benchmarking should be used for

<table>
<thead>
<tr>
<th></th>
<th>3.37</th>
<th>0.88</th>
<th>0.00</th>
<th>51.43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum numbers of operative cases trainees should perform.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum proportions of operative cases trainees should perform.</td>
<td>3.60</td>
<td>0.69</td>
<td>0.00</td>
<td>60.00</td>
</tr>
</tbody>
</table>

### Trainee improvement

<table>
<thead>
<tr>
<th></th>
<th>3.76</th>
<th>0.73</th>
<th>0.03</th>
<th>70.59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative assessment should be in the form of a written assessment of trainees.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### How should the results of measuring surgical training quality be used?

<table>
<thead>
<tr>
<th></th>
<th>3.74</th>
<th>1.01</th>
<th>0.00</th>
<th>65.71</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transparency: Should be made freely available.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Measuring the quality of surgical training will:

<table>
<thead>
<tr>
<th></th>
<th>3.86</th>
<th>0.91</th>
<th>0.00</th>
<th>68.57</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create accountability for Hospital trusts.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Measuring the Quality of Surgical Training

- Operating Room
  - Trainees should be involved in selection of cases that are placed on a theatre list.
  
  | 3.34 | 0.18 | 0.00 | 54.29 |

### Measuring the Quality of Surgical Training

- Logbook analysis
  - Trainee logbook analysis is essential to measure surgical training quality.
  
  | 3.77 | 0.84 | 0.00 | 68.57 |

  - Aggregate measures of trainee logbooks should be used.
    
    | 3.51 | 0.70 | 0.00 | 51.43 |

  - Trainer logbook analysis is essential to measure surgical training quality: case load versus how much their trainees do.
    
    | 3.66 | 0.64 | 0.00 | 57.14 |

  - Logbooks should include clinic numbers in addition to operative cases.
    
    | 3.54 | 0.82 | 0.00 | 57.14 |

  - Logbooks should be modified to allow greater detail about how much of the operation a trainee has performed.
    
    | 3.77 | 0.81 | 0.00 | 65.71 |

- The effect of timetable/rota structure on:
  - The continued interaction between the trainee and their trainer should be assessed.
    
    | 3.71 | 0.71 | 0.00 | 68.57 |

  - The continuity of care for patients should be assessed.
    
    | 3.83 | 0.79 | 0.00 | 71.43 |

- Multi-source feedback to measure surgical training quality?
  - Independent (third party) feedback is an essential component.
    
    | 3.37 | 0.91 | 0.00 | 51.43 |

- Data on training quality should be sent to a central administration unit to collate and process the results.
    
    | 3.47 | 0.95 | 0.03 | 50.00 |

- Objective numbers such as case load can be provided by hospital administrators or national databases such as Hospital Episode Statistics.
    
    | 3.37 | 0.88 | 0.00 | 51.43 |

- Benchmarking should be used for:
  - Minimum numbers of operative cases trainees should perform.
    
    | 3.63 | 0.69 | 0.00 | 62.86 |

  - Minimum proportions of operative cases trainees should perform.
    
    | 3.60 | 0.69 | 0.00 | 60.00 |

- Trainee improvement
  - Formative assessment should be in the form of a written assessment of trainees.
    
    | 3.76 | 0.73 | 0.03 | 70.59 |

- How should the results of measuring surgical training quality be used?
  - Transparency: Should be made freely available.
    
    | 3.74 | 1.01 | 0.00 | 65.71 |

- Measuring the quality of surgical training will:
  - Create accountability for Hospital trusts.
    
    | 3.86 | 0.91 | 0.00 | 68.57 |
Appendix D: Sample SIMweek OSCE Score Sheet

**Arterial Blood Gas**

**Generic requirements:**
- introduce yourself
- check the patient’s identity: name and hospital number on I.D.
- confirm that the procedure is required
- explain the procedure to the patient (including possible complications and risks)
- gain informed consent for the procedure (under direct supervision where appropriate)
- take all necessary steps to reduce the risk of infection,
- assembles correct equipment in the tray
  - including washing hands,
  - wearing gloves and maintaining a sterile field if appropriate
- dispose of all equipment in the appropriate receptacles
- document the procedure in the notes; and
- arrange appropriate aftercare/monitoring.

**Procedure specific requirements:**
- prepare Arterial Blood Gas (ABG) syringe, skin cleaning material
- check expiry date and expel Heparin (if required to)
- clean and palpate artery with index and middle fingers
- insert needle between fingers at 45 degree angle until blood enters syringe.
  - Arterial pressure will usually fill the syringe
- withdraw and ask assistant to apply pressure via cotton wool ball for five minutes
- apply filter to syringe, hold upright and expel air
- roll to mix, confirm label and send to lab.

**GLOBAL MARK**
- 5 very good
- 4 good
- 3 pass
- 2 borderline
- 1 fail.
Appendix E: SIMweek Confidence Questionnaire

Please rate your confidence levels to manage the following conditions or perform the following procedures:

1. Initial management of patients with peri-operative chest pain
   (No confidence) 1 2 3 4 5 (Confident)

2. Initial management of a post-operative fever
   (No confidence) 1 2 3 4 5 (Confident)

3. Initial management of sepsis
   (No confidence) 1 2 3 4 5 (Confident)

4. Initial management of post-operative bleeding
   (No confidence) 1 2 3 4 5 (Confident)

5. Acute management of a GI bleed
   (No confidence) 1 2 3 4 5 (Confident)

6. Initial management of oliguria
   (No confidence) 1 2 3 4 5 (Confident)

7. Placing a peripheral IV cannula
   (No confidence) 1 2 3 4 5 (Confident)

8. Placing a Foley urinary catheter in a female patient
   (No confidence) 1 2 3 4 5 (Confident)

9. Placing a Foley urinary catheter in a male patient
   (No confidence) 1 2 3 4 5 (Confident)

10. Managing electrolyte imbalances in perioperative patients
    (No confidence) 1 2 3 4 5 (Confident)

11. Prescribing fluids and electrolytes
    (No confidence) 1 2 3 4 5 (Confident)

12. Nasogastric tube placement and management
    (No confidence) 1 2 3 4 5 (Confident)

13. Basic peri-operative pain management
    (No confidence) 1 2 3 4 5 (Confident)

14. Write and understand the essential components of a daily progress note
    (No confidence) 1 2 3 4 5 (Confident)
15. Understand the essential components of a surgical note  
(No confidence) 1 2 3 4 5 (Confident)  

16. Prescribing common medications that surgical FY1s order  
(No confidence) 1 2 3 4 5 (Confident)  

17. Identify correct central venous catheter placement on a chest radiograph  
(No confidence) 1 2 3 4 5 (Confident)  

18. Identify correct nasogastric tube placement on a radiograph  
(No confidence) 1 2 3 4 5 (Confident)  

19. Identify common surgical diagnoses on differing radiographic studies  
(No confidence) 1 2 3 4 5 (Confident)  

20. How to prep and drape a surgical patient  
(No confidence) 1 2 3 4 5 (Confident)  

21. Perform an instrument tie  
(No confidence) 1 2 3 4 5 (Confident)  

22. Perform a simple interrupted closure  
(No confidence) 1 2 3 4 5 (Confident)  

23. Place a horizontal mattress suture  
(No confidence) 1 2 3 4 5 (Confident)  

24. Perform a vertical mattress suture  
(No confidence) 1 2 3 4 5 (Confident)  

25. Make a skin incision  
(No confidence) 1 2 3 4 5 (Confident)  

26. Excise a 2-cm sebaceous cyst on the back  
(No confidence) 1 2 3 4 5 (Confident)  

27. Know basic settings on lap camera/insufflator  
(No confidence) 1 2 3 4 5 (Confident)  

28. Drive a 0-degree laparoscopic camera  
(No confidence) 1 2 3 4 5 (Confident)  

29. Drive a 30-degree laparoscopic camera  
(No confidence) 1 2 3 4 5 (Confident)  

30. Identifying surgical instruments by name and function  
(No confidence) 1 2 3 4 5 (Confident)
31. Perform and interpret arterial blood gas analysis
   (No confidence) 1 2 3 4 5 (Confident)

32. Identify anatomical images on CT Scan
   (No confidence) 1 2 3 4 5 (Confident)

33. Identify anatomical images on radiographs
   (No confidence) 1 2 3 4 5 (Confident)

34. Global confidence
   (No confidence) 1 2 3 4 5 (Confident)

35. I am/was well prepared to be a surgical FY1 at the start of my job
   (No confidence) 1 2 3 4 5 (Confident)