Exploring surgical and clinical skills learning in postgraduate and undergraduates

Alexander W Phillips

A thesis in fulfilment of the degree of Doctor of Medicine

(By Publication)

Faculty of Medicine and Health Sciences

University of East Anglia

March 2017

Abstract

There has been a huge change in both undergraduate and postgraduate training over the last 20 years. This has been due to a number of issues, including a desire for greater transparency, an emphasis on quality of training, and a desire to streamline training with a consequent drop in training hours available and a need to optimise every training opportunity to the maximum. This has led to an increased emphasis on quality of teaching and a recognition of the importance of timely and pertinent feedback. Within surgery there has been the creation of a new curriculum. This places competency based training at its centre and incorporates Workplace Based Assessments (WBAs) as a key component of proving competency. Their use has been accepted with some misgivings.

Further, the reduction in training hours has led to a need to supplement hours spent "working" with alternative training mechanisms. Simulation has been increasingly favoured, but much work needs to be done to ensure appropriate training and appropriate feedback can be obtained from simulators.

Both quantitative and qualitative research strategies have been employed to determine the changes in training that have occurred, their impact and how stakeholders can view the implementation of different feedback mechanisms.

The findings have demonstrated an appreciation for the use of WBAs, albeit with a number of reservations including concerns regarding validity and reliability and sufficient opportunity for their completion. The papers included have also demonstrated that effective feedback from simulated training is possible and that self-evaluation is as effective as having expert tuition for simple tasks.

Training for technical skills has evolved from the historic "see one, do one, teach one" model and now requires high-quality and validated training and feedback. Studies into how best this can be provided will continue to change as the learning environment changes.

For my mum, who has supported everything I do.

Contents

List of Figures

Abbreviations

ASIT	Association of Surgeons in Training
ARCP	Annual Review of Competence Progression
CBD	Case Based Discussion
CCT	Certificate of Completion of Training
CCST	Certificate of Completion of Specialist Training
CST	Core Surgical Trainee
EMQ	Extended Matching Question
EWTD	European Working Time Directive
FRCS	Fellowship of the Royal College of Surgeons
GMC	General Medical Council
GP	General Practitioner
HST	Higher Surgical Trainee
ISCP	Intercollegiate Surgical Curriculum Programme
Mini-CEX	Mini-Clinical Evaluation Exercise
MRCS	Membership of the Royal College of Surgeons
MMC	Modernising Medical Careers
MCQ	Multiple Choice Questions
OSCE	Objective Structured Clinical Examination
PBA	Procedural Based Assessment
PMETB	Postgraduate Medical Education Board
PRHO	Pre Registration House Officer
RITA	Record of In-Training Assessment
RCS	Royal College of Surgeons
SHO	Senior House Officer
WBA	Workplace Based Assessments

Acknowledgements

I would first like to thank my co-authors on the papers presented in this thesis. They have been a pleasure to collaborate with and their efforts and encouragement have led to the success of each of these studies.

Second, there is a debt of gratitude to those students and trainers who signed up to take part in trials and answered seemingly endless questions, and who gave their time selflessly. I hope they learnt something, and I hope that some of the work included in this study, through their efforts, will go on to improve training for others.

Third, my gratitude to Mr James Hernon and Professor Sam Leinster who supervised this thesis.

Finally, I am grateful to the ongoing support of my family and friends without whom I would never have sought to complete this.

Chapter 1: Introduction

1.1 Introduction

Medical training has changed drastically over the last 20 years. These changes have affected both undergraduate and postgraduate curricula, and are widespread through every subspecialty. There is an ever greater emphasis on the quality of teaching provided, and on the provision of timely and appropriate feedback.

The integration of Workplace Based Assessments (WBAs) has been a fundamental method of trying to ensure feedback is given, as well as providing better quality of feedback. The assessments, additionally, are integrated into a portfolio of evidence of progression. They have become increasingly important as the time in training has declined due to a number of factors, including the European Working Time Directive (EWTD), and a drive towards streamlining training.

Whilst feedback has been recognised as an important component of learning, there has been little evidence for these WBAs. In addition, there is increasing use of technology and simulation to aid learning to compliment and supplement the clinical opportunities that are available.

This thesis, through a collection of published works, explores the progression of the surgical curriculum, the use of WBAs that make up a fundamental component of the new curriculum, and the use of simulation for learning simple technical skills.

1.2 Thesis Structure

This thesis explores the learning of technical skills in both undergraduate and postgraduate clinicians and the use of WBAs to aid learning.

The portfolio of papers included within this thesis is presented along three key themes based on surgical training and formative feedback received in the UK. This work has made original contributions to a variety of key areas in this field including:

- The evolution of surgical training and presentation of training experience in a technically demanding procedure (oesophagectomy) highlighting that this does not compromise outcomes.
- A review of the validity and reliability of commonly used WBAs in surgical training.
- The impact of self-review of videos on learning technical skills.
- Perceptions of both trainees and trainers on WBAs in surgical training.

The first chapter outlines the changes that have occurred in surgical training. Two papers are included within this chapter. The first draws direct comparison between the current surgical curriculum and its predecessor. The second paper review outcomes of training within this new curriculum for a technically advanced procedure and establishes that training can occur for such procedures with no detriment to patient care.

The second chapter looks specifically at feedback; first with regard to the validity and reliability of WBAs which have become a core component of medical training, and secondly with regard to feedback received in simulated learning of simple practical skills. Simulation within training is discussed given that this is a core aspect of the two trials that are included.

The third chapter looks at perceptions of feedback mechanisms and discusses WBAs from the point of view of the assessor. A qualitative study into a specific type of WBA (the Case Based Discussion) is included, and the methodology is discussed at length given the limitations of the manuscript that was published in *Clinical Teacher*. Finally, a paper reviewing the feedback mechanisms within the two trials from the previous chapter is included to further evaluate the feedback mechanisms used within the trials.

A discussion is included at the end of each chapter of the papers supplied within that theme and a critical appraisal of the papers within each chapter.

1.3 Thesis Objectives

The key objectives for the work undertaken as part of this thesis were to:

1. Analyse the changes that have occurred in surgical training and the reasons behind them.

2. Investigate if high-quality training can occur within the current surgical training programme.
3. Review the validity and reliability of WBAs within surgical training.
4. Investigate the impact of video-based feedback on learning technical skills.
5. Explore the perceptions of users of feedback mechanisms.

Collectively these objectives aim to improve training by determining how training occurs and by identifying the tools that may potentially lead to better learning.

1.4 List of Published Works Submitted for Thesis

The publications upon which this thesis is based are included within the body of this thesis and include:

Theme 1: Background to Surgical Training

1. Phillips AW, Madhavan A. A critical evaluation of the Intercollegiate Surgical Curriculum and comparison with its predecessor the "Calman" curriculum. *J Surg Education* 2013. 70(5): 557-562[1]
2. Phillips AW, Dent B, Navidi M, Immanuel A, Griffin SM. Trainee involvement in Ivor Lewis Esophagectomy does not negatively impact outcomes. *Annals of Surgery*. 2018; 267(1):94-98[2]

Theme 2: Performance assessment and feedback of simple clinical skills

3. Phillips AW, Jones AE. The validity and reliability of workplace-based assessments in surgical training. *Bulletin of The Royal College of Surgeons of England* 03/2015; 97(3):e19-e23[3]
4. Nesbitt CI, Phillips AW, Searle RF, Stansby G. Randomised trial to assess the impact of supervised and unsupervised video feedback on teaching practical skills. *J Surg Educ*. 2015 Jul-Aug;72(4):697-703[4]

5. Phillips AW, Matthan J, Bookless LR Whitehead IJ, Madhavan A, Rodham P, Porter AL, Nesbitt CI, Stansby G. Individualised expert feedback is not essential for improving clinical skills performance in novice learners: A randomised trial. *J Surg Ed.* 2017 Jul-Aug; 74(4): 612-620 [5]

Theme 3: Perceptions of formative assessment

6. Phillips AW, Madhavan A, Bookless LR, Macafee DA. Surgical Trainers' Experience and Perspectives on Workplace Based Assessments. *J Surg Educ*. 2015 Sep-Oct;72(5):979-84[6]
7. Phillips A, Lim J, Madhavan A, Macafee D. Case-based discussions: UK Surgical Trainee perceptions. *Clin Teach*. 2016 13(3): 207-129[7]
8. Nesbitt CI, Phillips AW, Searle RF, Stansby G. Student views on the use of two styles of video-enhanced feedback compared to standard lecture feedback during clinical skills training. *J Surg Educ*. 2015 Sep-Oct;72(5):969-73[8]

1.5 Declaration of Published Work

Below are listed each of the works included as part of the thesis and my contribution to them. In every piece of work I wrote the first draft of the manuscript, and completed revisions.

1. Phillips AW, Madhavan A. A critical evaluation of the Intercollegiate Surgical Curriculum and comparison with its predecessor the "Calman" curriculum. *J Surg Education* 2013. 70(5): 557-562

   I developed the question and carried out a literature review relating to this paper. I carried out the comparison of the two curricula and wrote the first draft. I was responsible for the submission and am the corresponding author.

2. Phillips AW, Dent B, Navidi M, Immanuel A, Griffin SM. Trainee involvement in Ivor Lewis Esophagectomy does not negatively impact on outcomes. *Annals of Surgery* 2018; 267:94-98

   I came up with the research question in this study, wrote the initial protocol and compiled and completed the patient database. I was responsible for completing the literature review, analysing the data and writing the first draft for this paper. I am the corresponding author.

3. Phillips AW, Jones AE. The validity and reliability of workplace-based assessments in surgical training. *Bulletin of The Royal College of Surgeons of England* 03/2015; 97(3):e19-e23

   I developed the research question and carried out a literature review relating to this paper. I carried out the analysis and wrote the first draft. I was responsible for the submission and am the corresponding author.

4. Nesbitt CI, Phillips AW, Searle RF, Stansby G. Randomised trial to assess the impact of supervised and unsupervised video feedback on teaching practical skills. *J Surg Educ*. 2015 Jul-Aug;72(4):697-703

   I helped develop the research question in collaboration with CI Nesbitt and G Stansby. I wrote the first draft of the manuscript and responded to the reviewers' comments.

5. Phillips AW, Bookless LR, Matthan J, Whitehead I, Madhavan A, Nesbitt CI, Stansby G. Individualised expert feedback is not essential for improving clinical skills performance in novice learners: A randomised trial. *J Surg Ed* 2017 Jul-Aug; 74(4):612-620

   I was responsible for writing the protocol and submitting for ethical approval. I analysed the results, and wrote the first draft. I submitted the paper for publication, responded to reviewers' comments and am the corresponding author.

6. Phillips AW. Madhavan A, Bookless LR, Macafee DA. Surgical Trainers' Experience and Perspectives on Workplace Based Assessments. *J Surg Educ*. 2015 Sep-Oct;72(5):979-84.

   I was responsible for the original concept designing the questionnaire and analysing the results. I wrote the initial draft was responsible for addressing reviewers' comments and am the corresponding author.

7. Phillips A**,** Lim J, Madhavan A, Macafee D. Cased-based discussions: UK Surgical Trainee Perceptions. *Clin Teach*. 2016;13:207-212

   I was responsible for writing the protocol and submitting for ethical approval. I analysed the results, and wrote the first draft. I submitted the paper for publication, responded to reviewers' comments and am the corresponding author.

8. Nesbitt CI, Phillips AW, Searle RF, Stansby G. Student views on the use of two styles of video-enhanced feedback compared to standard lecture feedback during clinical skills training. J Surg Educ. 2015 Sep-Oct;72(5):969-73

   I was responsible for conceiving the study, analysed the data wrote the initial draft and responded to the reviewers' comments.

1.6 Declaration of Contribution

Declaration regarding co-authorship:

Titles:

1) Nesbitt CI, Phillips AW, Searle RF, Stansby G. Randomised trial to assess the impact of supervised and unsupervised video feedback on teaching practical skills. J Surg Educ. 2015 Jul-Aug;72(4):697-703

2) Nesbitt CI, Phillips AW, Searle RF, Stansby G. Student views on the use of two styles of video feedback compared to standard lecture feedback. J Surg Educ. 2015 Sep-Oct;72(5):969-73

Confirmation of contribution of AW Phillips by first author:

I Craig Nesbitt confirm that Alexander Phillips made a significant contribution to both of these papers. He was involved in formulating the original plan and methods of the study, analysis of the data in preparation for the manuscript and writing the initial draft and revisions of each manuscript.

Signature.....

Chapter 2: Background, Context and Rationalisation

Theme 1: Background to Surgical Training

1. Phillips AW, Madhavan A. A critical evaluation of the Intercollegiate Surgical Curriculum and comparison with its predecessor the "Calman" curriculum. *J Surg Education* 2013. 70(5): 557-562[1]
2. Phillips AW, Dent B, Navidi M, Immanuel A, Griffin SM. Trainee involvement in Ivor Lewis Esophagectomy does not negatively impact outcomes. *Annals of Surgery* 2018; 267: 94-98[2]

This chapter looks at the current state of surgical skills training and provides a background to the fundamental changes that have occurred within the surgical curriculum. The first paper evaluates the changes between the old curriculum and the new curriculum. The second paper evaluates training for a specific operation (oesophagectomy) during the new curriculum.

2.1 Background to Surgical Training

2.1.1 How Surgical Training has Evolved

Training in surgery has traditionally followed an apprenticeship model. William Halstead, an American Surgeon who pioneered the use of the aseptic technique, coined the phrase “See one, do one, teach one” a mantra that has been repeated by trainers frequently over the last

century[9]. This ethos of trainees observing a procedure before actually performing the skill and then consolidating their learning by teaching has become contentious given the risk to patient safety, which is paramount[10–12]. There are a number of factors impacting upon current surgical training including patient safety, new technologies, changing working roles, the balance of training and service provision and a reduction in actual time for learning[13–15]. These factors have all served to drive huge change in the way a surgeon is trained[16]. The reasons for a decline in actual time for training, which has been apparent globally[16–18], are multi-factorial and includes reduced working hours (as determined by the EWTD), a drive for surgeries to be led by Consultants, and increasing subspecialisation. These have all contributed to a fall in actual operative numbers for those in training.

Historically, within the UK, there was no defined standard that trainees had to attain before being regarded as competent[1]. Evidence was usually in the form of a substantial logbook which would display the length and breadth of experience that a trainee had achieved. This has all been transformed in the last ten years with the emphasis on competency-based training and assessment. This paradigm shift aimed to place emphasis on evidence of skill attainment rather than the number of hours spent training, or the number of procedures performed.

The use of operative volume as an indicator of quality has been supported by evidence that patient outcomes improve with increased operative volume[19]. This volume-outcome relationship which has been demonstrated in both individual surgeons and in hospital volume[20] has led to centralisation of complex procedures and consequent increasing subspecialisation. Centralisation of procedures can have a positive and negative impact on the trainee. For those working in high-volume specialist units it allows experience to be concentrated and skills to be consolidated by repetition. However, other trainees may be affected by the lack of exposure and lack of opportunity to work in such units. One consequence of centralisation is that trainees will need to work in high volume specialist units in order to obtain the necessary experience. This may potentially take the form of an additional year's fellowship.

Defining competence has been a key theme to the changing surgical curriculum. The previous incarnation of surgical training known as the "Calman" curriculum, involved trainees participating in a higher surgical training programme which usually followed on from a pre-registration (or intern year) and then a period as a junior surgical trainee (also known as a Senior House Officer SHO)[1]. Centralisation of complex and less frequently performed procedures has led to trainees undertaking fellowships. In the UK the Association of

Surgeons in Training (ASIT) conducted a survey on UK trainees and found over 75% intended to undertake fellowships on "completion" of their training[21]. These findings have been echoed in the US where more than 80% of trainees plan fellowships[16].

2.1.2 Creation of the Calman Curriculum

The eponymous Calman reforms were named after the Chief Medical Officer between 1991 and 1997, Sir Kenneth Calman. His proposed changes published within "*Hospital Doctors-Training for the Future*"[22], aimed to improve training and provide a better structure for hospital trainees. The reforms focused on those in the latter parts of their training and largely did not affect the more junior Pre-Registration House Officer (PRHO) and SHO grades. One component of the reforms was the replacement of the Registrar and Senior Registrar training posts with the Specialist Registrar grade. This change was implemented in order to offset the build-up of a large number of experienced doctors within the Senior Registrar grade who were waiting for a consultant post. The changes brought about reduction in time spent training with the aim of reducing the average training duration from 12 years to approximately 7 years (See Figure 1- *Training Structure*). With the more defined timeframe and better structure to the training programme came a recognised end point and the award of a Certificate of Completion of Training (CCT) awarded by the General Medical Council (GMC).

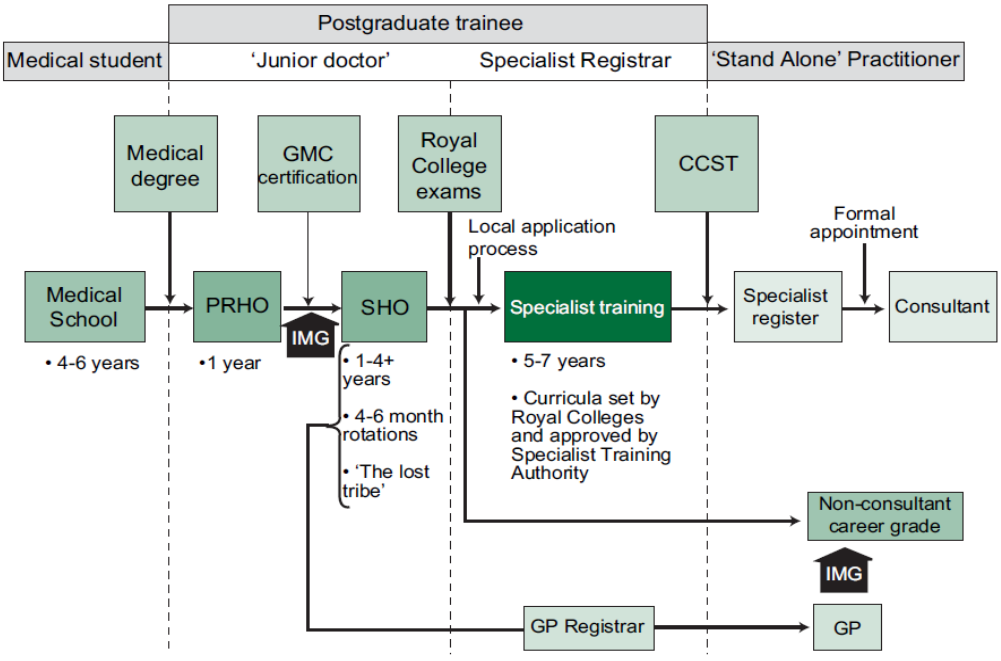

Figure 1: UK training following the Calman reforms[23]

As well as differing from its predecessor by shortening the time spent within the training programme, the relatively "new" specialist registrar grade included a number of other changes. Competitive entry with defined pre-requisite experience was stipulated and trainees were to have an annual review in the form of a Record of In-Training Assessment (RITA), to gauge progress and permit progression to the next year of training.

Interestingly, with these reforms came the concerns that this shortened training would lead to consultants not having the knowledge or experience to practice safely[24].

2.1.3 Senior House Officer Reform.

The turn of the millennium saw another Chief Medical Officer, Sir Liam Donaldson, publish his own treatise on how to deal with the SHO grade in "*Unfinished Business- Proposals for reform of the Senior House Officer grade*"[25]. Whilst the Calman reforms had sought to streamline higher specialist training and effectively ensured that appointment to a registrar training number should lead to appointment as a consultant, the "core" group of trainees immediately following on from their PRHO year had to some extent been isolated. Trainees would usually undergo their initial training in whichever subspecialty interested them following their PRHO year. However, no real career structure existed for these doctors, and many trainees could work for indefinite periods at SHO level with little career development. Whilst some could obtain important experience, no credit at a higher level could be attained. Many jobs were short, six month posts, which was unsettling and necessitated relocation around the country. Those posts that were rotational did not have a structure or curriculum aimed at providing a standardised level of experience.

With no clear provision of goals, and no defined competencies, the *curriculum* issued by the Royal College of Surgeons (RCS) was actually no more than a syllabus. It listed components but did not provide any real detail as to levels of knowledge and clinical skills, and how they should be obtained. The only objective discriminator that existed was the RCS examinations and it was upon these solely that trainees were judged.

The changes to higher specialist training served to create a bottleneck in the training pathway between the SHO level and the registrar grade. The lack of a defined end-point to SHO training meant that a large cohort of trainees became stranded at this level through failure to

secure promotion to specialist training programmes. This was reflected by evidence from a report that indicated that half of NHS trainees were within the SHO grade[25].

Donaldson's report produced five key recommendations which led to the reformation of all medical training within the UK. These were that:

1) Training should be programme based.

2) Training should begin with broadly-based programmes pursued by all trainees.

3) Programmes should be time-limited.

4) Training should allow for individually tailored or personal programmes.

5) Arrangements should facilitate movement into and out of training and between training programmes.

The changes that were proposed involved changing the previous PRHO year into a two-year "Foundation Programme" (although trainees would be eligible for full GMC registration at the end of their first year), followed by commencement of specialist training. It was felt that such a Foundation Programme would provide trainees with the opportunity to develop a wide range of generic skills that could be transferred to whichever specialty they eventually ended up working in. It would also, theoretically, provide newly qualified doctors with an increased opportunity to experience multiple medical specialities and help them select the most appropriate specialty for their future career[25].

Specialist training programmes would follow this period of foundation training, and would follow the five key principles above. It was also realised that in order to provide efficient training, within the confines of the pending EWTD, changes to working patterns and efficiency in utilising training opportunities available would be vital.

The impact of changing the early years led to discussion on how these initial years of training should be linked into specialist training[26]. The concept of "run-through" training was introduced, which involved trainees applying for specialist training which would commence on completion of their foundation programme. This would mean doctors would need to make a single application that would then see them in training posts up until they reached consultant level. It was also hoped that this would streamline training, making it more efficient as it became competency based, and shorten the time taken to become a consultant so that it was more in line with European and US counterparts. It was appreciated that this

may produce consultants with a more generic skill set and that there may indeed be junior consultants who were then required undertake further subspecialty training. These proposals did have the significant disadvantage of removing flexibility- one of the key principles mooted in the report.

Amongst the various reforms that were to be implemented was the dependence on a competency-based system. The importance of evidencing skills acquired was emphasised and trainees would be required to have regular in-programme assessments and demonstrate progress by use of a multitude of WBAs, which provided formative feedback.

2.1.4 Modernising Medical Careers

The proposed reforms were regarded as the way forward and the Department of Health produced a policy on "*Modernising Medical Careers*" in 2003[23]. A further publication "*Modernising Medical Careers: The Next steps*" was published in 2004. This report detailed how changes in doctor training were to be taken forward[26] (figure 2).

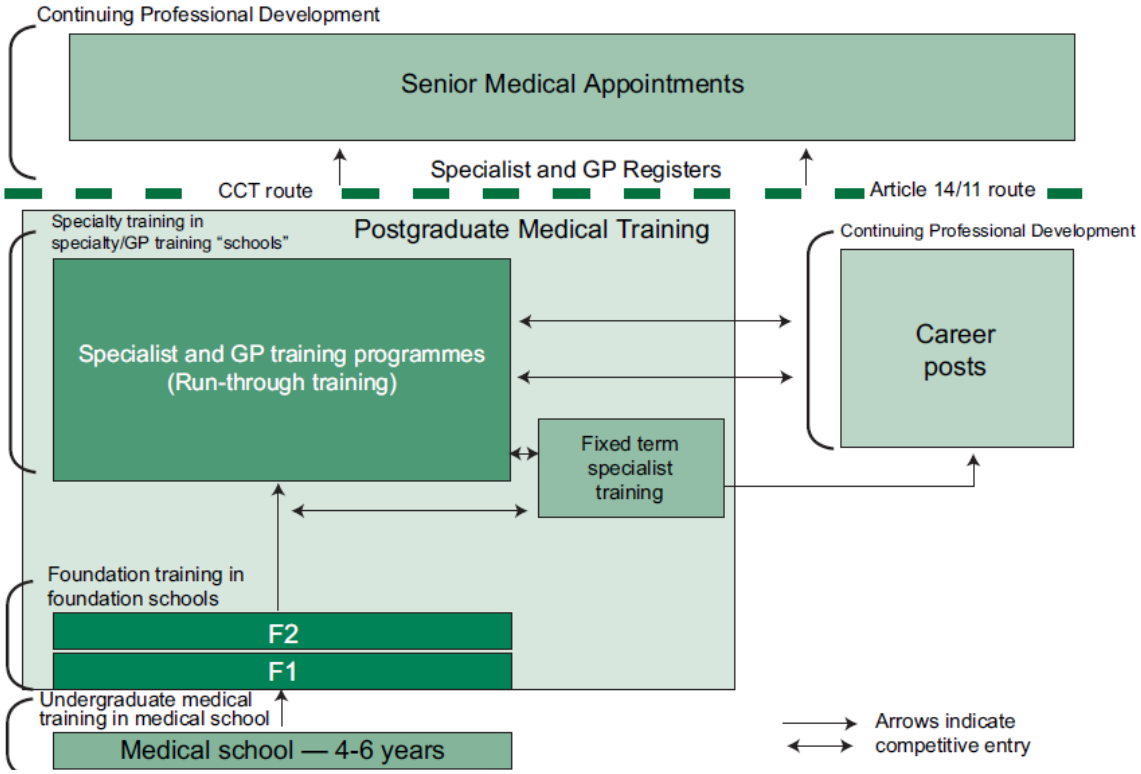

Figure 2: Structure of medical training proposed by MMC[23]

The principles that were part of foundation–level changes were incorporated and included the aims that it should be:

1) Trainee-centred

2) Competency-assessed

3) Service based

4) Quality assured

5) Flexible

6) Coached

7) Structured and streamlined.

Foundation programme pilots commenced in 2004. The "*Unfinished Business*" paper introduced the possibility of "run-through" training which was eventually adopted across many specialties including surgery[25]. This would mean a single entry point into specialist training once doctors had completed their Foundation years and met the competencies held within that programme. Once selected for specialist training, a doctor would have no further competitive interviews until consultant applications provided they met the competencies within their training programme.

2.1.5 Reasons for Training Reform

There were a number of reasons for seeking these reforms, many of which have already been alluded to. One of the major reasons was the balance that needed to be met between training of junior doctors and service provision. There have been increasing demands and emphasis on the latter due to a shortage of doctors and also an increased drive to maximise efficiency within the NHS[27]. For junior doctors, increased efficiency is partly dependent on the experience they have obtained whilst training. Fewer working hours through curriculum reform has seen a change in how junior doctors work and has led to a fall in weekly working hours. For surgical trainees, this has been postulated as a two-thirds fall in hours worked. The majority of trainees have suggested that the changes to their training have been detrimental, and that the new working patterns do not allow for sufficient experience to be gained. Thus, the expectation has become that junior doctors can achieve the requisite standards to practice

as consultants despite fewer hours in which to do so. To offset this an increased emphasis in improving teaching standards and having a more prescriptive curriculum was required[27–29] .

There already exists a great deal of variability in the quality of teaching and assessment throughout the undergraduate and postgraduate years. The Quality Assurance Agency (1998-2002) assessments reviewed teaching standards across UK medical schools, revealing a disparity with some medical schools achieving excellence and others struggling[30]. This was compounded by the drive to increase medical student numbers with a 60% increase in the number of students between 1998 and 2003. A further 1500 places were announced in October 2016 to be included for the academic year 2018/19 as a drive occurred to make the NHS more self-sufficient in the provision of doctors[31]. This contributes towards an imbalance between educational demand and resource availability. Ongoing reduction in working hours, legislated for by the EWTD, and staff shortages mean that there is a constant pressure on being able to provide small group teaching programmes. This is due to the lack of a sufficient number of educators[32].

There is also a great variability in teaching standards in the postgraduate setting. It has been suggested that this is due to a lack of training for clinical teachers[33]. Most health service staff are short of time and have limited training in teaching methods[34]. This coupled with the increasing number of medical undergraduate students[32,35,36] can jeopardise the quality of teaching. One study demonstrated over three-quarters of clinical teachers had received no formal training[37] despite a drive for clinical excellence through clinical governance, consultant appraisal and revalidation. Consultants are obliged to have an annual appraisal which emphasises teaching and receiving feedback as a measurement of its success[38,39].

In a separate paper to his review of specialist training, Calman felt that the way to address the diversity of quality was "*to look at how doctors learn and how teaching methods can be improved in order that the time spent in education can be used efficiently*". He wrote, "*The integration of theoretical teaching with practical work, progressive assessment and feedback to teachers and trainees are essential*"[40].

This ethos was echoed in the US by the American Committee of Deans[41] and also by the Committee of Postgraduate Deans within the UK[42]. Thus a feeling existed that clinical teaching and training needed to be improved so that the maximum benefit could be gleaned from them. Teaching needed to go hand-in-hand with service provision and activities often

seen as service provision i.e. outpatient clinics, ward rounds, needed to be used as learning opportunities[43].

2.2 Learning within the Clinical Environment

Learning within the clinical environment is distinctly different from what many students will have encountered previously as either lectures or seminars. Teaching within the clinical setting involves the use of practical skills, whether it be communication skills or performing procedures on patients. Kolb in his "experiential learning theory" suggested that the "concrete experience" that was obtained from these practical experiences enabled students to reflect on their practice (reflective observation) and to try and determine the underlying reason for what ensued (abstract conceptualisation). Following this pathway students should be able to apply what they have learnt to further problems (active experimentation) to complete the learning cycle[44](figure 3).

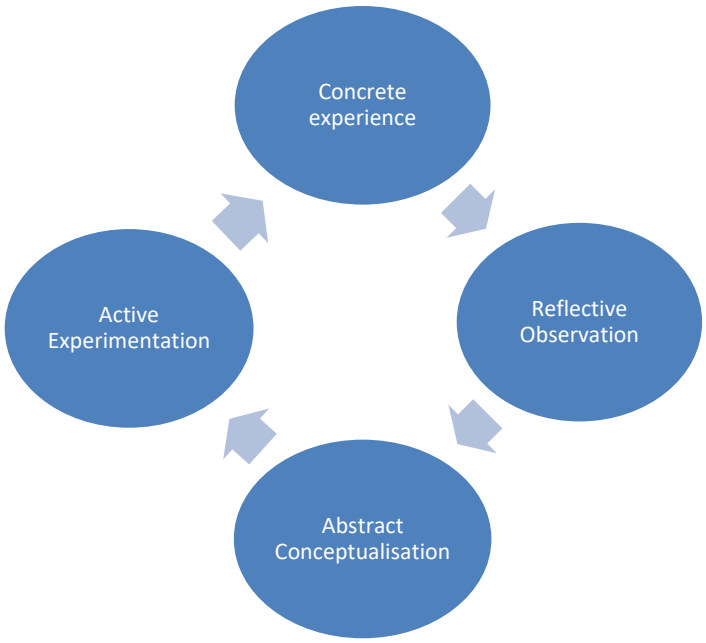

Figure 3: Kolb's Experiential Learning Cycle[44]

This learning cycle can be used to benefit learners by giving them the opportunity for 'practice' and providing immediate feedback. Following this, the learner may review, and think of alternative strategies helping to generalize learning. Campbell et al[45] carried out a

study based on Schon's 'reflective practitioner'[46] where doctors were asked to keep reflective diaries. Doctors wrote down any event concerning their clinical practice that stimulated their thinking and which they would then reflect on ('Reflection on Action'). They found that the events that provided the greatest stimulus to learning were reading literature and patient management. Schon's theory incorporates a key element of self-directed learning and the results suggested that clinical experience with patients is important to this.

Not all clinical learning opportunities are however positive[47]. It has been suggested that there is a close and vital relationship between experience and education and that some experiences can be detrimental to learning[48]. What may prove beneficial for one trainee, could possibly be detrimental to another. Trainees have two competing requirements, first their need to learn, and second the service they are expected to provide within the healthcare system. It is possible that these may detract from each other restricting learning and reducing the important components of reflecting which would allow them to plan for the future.

2.3 Lifelong Learning

Parsell and Bligh indicated the importance of life-long learning: "*Medical education may be viewed as a learning continuum that only ceases at the end of the doctors' professional lives*"[49]. For doctors and in particular those that are in training, the majority of these learning opportunities that present are in the clinical environment. Others have emphasised the importance of making the most of clinical opportunities for learning as "postgraduate medical education is based on clinical practice"[42]. However, the great changes that have occurred to medical training with increased pressures on time and opportunities mean that these clinical opportunities need to be supplemented in an optimal fashion so that trainees can not only learn new skills outside of the clinical environment, but also consolidate skills that have been learnt. The importance of self-evaluation and the ability to self-evaluate needs to be established at an early stage of a medical student's career. Trainees need to be able to develop their ability to reflect and learn if they are to be able to maximise the new non-clinical opportunities that are available. In the era of increasing technological complexity and information availability, the skills of being able to identify learning goals and translate these goals into learning events are increasingly important for medical practitioners[50]. It is now impossible for doctors to know, and keep up-to-date with rapidly changing specialities. With

the increasing volume of medical literature available it is vital to learn how to become a lifelong learner[51].

Changes to what is regarded as best practice in clinical education have impacted greatly on the way students are taught, and learn. It is likely that there will be further evolution particularly with the increasing use of simulation. Such changes will affect the management of risk, trainee satisfaction and career choices[52–54].

2.4 Comparison of the Intercollegiate Surgical Curriculum and the Calman Curriculum: Paper and Commentary

1. Phillips AW, Madhavan A. A critical evaluation of the intercollegiate surgical curriculum and comparison with its predecessor the "Calman" curriculum. *J Surg Education* 2013. 70(5): 557-562[1]

A Critical Evaluation of the Intercollegiate Surgical Curriculum and Comparison With its Predecessor the "Calman" Curriculum

Alexander W. Phillips, MRCSEd, and Anantha Madhavan, MRCS

Department of General Surgery, James Cook University Hospital, Middlesbrough, United Kingdom

BACKGROUND: The increasing need for doctors to be accountable and an emphasis on competency have led to the evolution of medical curricula. The Intercollegiate Surgical Curriculum Project succeeded the Calman curriculum for surgical training in 2007 in the UK. It sought to provide an integrated curriculum based upon a website platform. The aim of this review is to examine the changes to the curriculum and effect on surgical training.

METHODS: A comparison was made of the Calman Curriculum and the ISCP and how they met training needs.

RESULTS: The new curriculum is multifaceted, providing a more prescriptive detail on what trainees should achieve and when, as well as allowing portfolio, learning agreements, and work-based assessments to be maintained on an easily accessed website. The increasing emphasis on work-based assessments has been one of the major components, with an aim of providing evidence of competence. However, there is dissatisfaction amongst trainees with this component which lacks convincing validity.

CONCLUSION: This new curriculum significantly differs from its predecessor which was essentially just a syllabus. It needs to continuously evolve to meet the needs of trainees whose training environment is ever changing. (J Surg 70:557-562. © 2013 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: surgical training, Calman curriculum, Intercollegiate Surgical Curriculum Program, work-based assessments

COMPETENCIES: Interpersonal and Communication Skills, Practice-Based Learning and Improvement, Professionalism

INTRODUCTION

The current edition of the surgical curriculum was implemented as part of the Modernizing Medical Careers changes. Review of the curriculum had been identified in the 1990s, and though previous reforms had sought to address higher surgical issues (Calman reforms), the new curriculum sought to integrate both basic and higher surgical trainee curricula to produce seamless training.

The aim here is to review the current intercollegiate surgical curriculum programme (ISCP)[1] in comparison with its predecessor, the Calman curriculum.[2]

The ISCP

Like most medical curricula, content is of paramount importance within surgery. The basis of its design must be to prepare, develop, and achieve a level of competence that allows one to practice independently as a consultant surgeon. The last 20 years have seen the demise of the true general surgeon with increasing subspecialization.[3] This is partly due to the rapidly expanding field of medical knowledge and skills (e.g., laparoscopic colorectal surgery) and research demonstrating that the centralization of certain surgical specialties leads to an improvement in patient outcomes.[4] This has resulted in a shift in what was deemed to be core knowledge and skills to specialty-specific competencies.[5]

PROGRAM RATIONALE

The ISCP was implemented in 2007 after 4 years of development. Revalidation and the 'Shipman' affair have also led to an emphasis on competency-based training and assessment and an encouragement of reflective practice.[6] (Harold Shipman was a British General Practitioner who was convicted as a serial killer after having murdered an

Correspondence: Inquiries to Alexander William Phillips, MA, MBBS, BSc, MRCSEd, James Cook University Hospital, Department of General Surgery, Marton Road, Middlesbrough TS4 3BW, UK; fax: +44 1642 282831; e-mail: awphillips@doctors.net.uk

Alexander W. Phillips came up with the original concept and prepared initial manuscript. Anantha Madhavan reviewed the manuscript and prepared for submission. Both authors reviewed the final manuscript.

Competing interests: Both authors are surgical trainees—Alexander W. Phillips is a higher surgical trainee and Anantha Madhavan is a core surgical trainee.

estimated 250 patients. He was undetected for many years having certified the deaths himself unchecked.) Four key concepts were identified as being necessary for a successful surgical curriculum:

(1) Focused training programmes underpinned by clear standards with defined levels of progression.
(2) Support to consultants to promote high-quality teaching and learning and reliable assessment.
(3) Rigorous and fully integrated regulatory systems, informed by curriculum standards.
(4) Adequate staff, resources, and reward systems to support trainees in attaining competence to Certificate of Completion of Training (CCT) level.[1]

The curriculum covers surgical training from post–foundation doctors through to consultant level, and is divided into 4 areas: syllabus, teaching and learning, assessment, and resources.

SYLLABUS

The syllabus provided includes an overview of the expected topics that should be covered by each trainee. These topics are further subdivided into more specific knowledge and skills. These are then graded on the level of competence that should be attained at each stage of training—initial, intermediate, and final based on the year of training. Competence ranges from 1—a basic level involving awareness of a fact of observation of a procedure, through to 4—knowing specific detail or showing competence at performing a procedure.

STAKEHOLDERS

Those involved in teaching and learning are carefully identified—program director (PD), assigned educational supervisors, clinical supervisors, assessors, and trainees. Each has a specific role in the implementation of the curriculum. Although the PD has overall responsibility for trainees in the deanery, the assigned educational supervisor is responsible for 1 to 4 trainees in their hospital, setting objectives, and ensuring satisfactory progress. The clinical supervisor is responsible for delivering, teaching, and ensuring that the trainees have sufficient resources available for them to meet their objectives. Assessors may come from a range of backgrounds and allow for continuing formative feedback. The final component is the trainees who have ultimate responsibility in driving their own training and ensuring they attain the appropriate levels of competence.

ASSESSMENT

The Membership of the Royal Colleges of Surgeons and Fellowship of the Royal Colleges of Surgeons examinations are the mainstay of summative assessment during the training. Work-based assessments (WBAs) are required to be carried out throughout training. These range from regular case-based discussions (CBDs), clinical evaluation exercises (CEXs), and directly observed procedures to procedure-based assessments (PBAs) at more advanced levels (Table 1). These then form a part of the summative annual review (Annual Review of Competence and Progression [ARCP]) along with the trainee's operative logbook. The ARCP also evaluates the trainee's academic achievements during their progress through the training program.

RESOURCES

Trainees and assessors are obliged to be registered to the web-based system (for which trainees pay an annual fee of £125). The website is a utility for documentation of learning plans and recording evidence of progress as well as outcomes of assessments. It also provides an online logbook and allows a portfolio of other activities to be documented.

TABLE 1. Types of Work Based Assessments and Their Purpose

Work-based Assessment	Domains Assessed	Key Features
Case-based discussion (CBD)	Communication, understanding, and application of knowledge	Evaluates patient management. Allows in-depth discussion to determine future learning needs.
Clinical evaluation exercise (CEX)	Clinical care skills	Assessment of a clinical encounter—provides feedback for history taking, examination, and communication.
Directly observed procedure (DOP)	Practical skills	Allows assessment against predefined checklist of steps in carrying out procedure.
Procedure-based assessment (PBA)	Practical skills	As with DOP—but more advanced procedures—i.e., in theater.
Multisource feedback (MSF)	Clinical care, maintaining good medical practice, learning and teaching, and professional relationships	Anonymized feedback from a variety of health professionals picked by trainee in a variety of work environments.

PROGRAM OUTCOMES

The final outcome for surgical training is judged on completion of the Fellowship of the Royal Colleges of Surgeons examination and a final ARCP assessment. However, satisfactory progress is necessary in each year of the training program, and this is assessed by completing a requisite number of WBAs (directly observed procedures, PBAs, mini-CEXs, and CBDs) and formation and satisfactory completion of learning agreements. There is an annual review of progress completed by a panel at deanery level which has superseded the old style record of in-training assessment. The curriculum allows flexibility in setting yearly outcomes and individualizing the curriculum depending on the long-term goals of the student.

QUALITY ASSURANCE

The responsibility of management and quality assurance lies with the schools of surgery within each deanery. The postgraduate deaneries (regionally based) must make certain that training posts are of requisite quality and assessments carried out within training posts are by trained assessors. In posts where training issues are raised, postgraduate deans have the power to trigger reviews of the training process. The deanery is thus the first line for ensuring that both trainers and trainees are maintaining expected levels. Ultimately, above deanery level, the General Medical Council (which merged in 2010 with the Postgraduate Medical Education and Training Board) assumes responsibility for entry, training, curriculum, and assessment standards that may comprise deanery visits and trainee surveys.[7]

Further, the Royal Colleges and specialist advisory committee have a duty to ensure a high standard of clinical care is maintained.[3]

COMPARISON WITH THE PREVIOUS CURRICULUM

Syllabus vs Curriculum

The new surgical curriculum has shown significant progress over its predecessor, the Calman curriculum. Indeed, it would appear that the Calman curriculum, to a large extent, regarded curriculum and syllabus as synonymous terms.

The new curriculum expands on a simple list of topics to be covered (a syllabus) and incorporates how, what, and when skills and knowledge should be acquired, as well as integrating a method of continuous formative assessment. The ISCP provides trainees with resources and formative assessments that continually evolve to meet their needs as they progress through their training. This provides a more structured and prescriptive source for learning. This has been a fundamental change in thinking rather than a gradual evolution away from an apprenticeship which is accompanied by learning topics.

Assessment vs Competency

Jolly and Rees described 4 basic types of medical curricula. These were content driven, method driven, assessment driven, and outcomes driven.[8] There has been a paradigm shift in surgical training away from a content-driven program that is identified by a syllabus to an outcome-based system that gives specific attainment targets at stages of training and an overall expectation a practitioner should be able to do at consultant level. This is particularly pertinent given the evolution toward subspecialization and a need to identify what *all* surgeons should be able to do vs what a subspecialty expert should be able to carry out.

The ISCP curriculum theoretically has moved itself away from being content driven by declaring that achieving competency is the important factor. It could be argued that to pass the Fellowship examination necessitates a level of competence; this cannot be said of the current Membership examination that exists as a summative hurdle before higher surgical training can be commenced. Paradoxically, training is not shortened in individuals that are able to prove competency more quickly, and achieving the requisite numbers of assessments in the work place has become the focus of determining satisfactory progress. Indeed the ISCP website is poorly regarded as a tool by trainees. Trainees were surveyed in 2008 with results indicating that 49% regarded the online assessments as poor or very poor and only 9% grading them good or very good.[9] A follow-up survey carried out in 2011 showed that there had been some improvement in how they were regarded, with 36% viewing them as poor compared with 22% regarding them as good.[10] There is also a concern that the assessments available lack validity.[11] This coupled with the technical problems and poor construction with regards to website navigability has meant that the new curriculum has not been well received.

There are a variety of reasons for shifting toward competency as the emphasis on training. There is a growing public demand that doctors must be accountable and must be able to demonstrate they have attained the required standard to practice autonomously as consultants.[12] Furthermore, the implementation of the European Working Time Directive would see a decline in operative exposure for surgical trainees. The original implementation of a 58-hour week in 2004 saw a 15.5% reduction in the number of procedures carried out by trainees.[13] The fall since implementation of a 48-hour week in August 2009 has yet to be properly determined. It is difficult to conclude

if the new curriculum has really taken account of the imposed time constraints. However, Parsons et al. demonstrated that independent operating ability of trainees had diminished significantly and few surgical trainees at Core Trainee level (the old senior house officer level) could perform appendicectomies (28%) or inguinal hernia repairs (8%) compared with current higher surgical trainees at the end of their core training (90% and 63%). There was also a general consensus amongst those interviewed at both core and higher surgical training levels that surgical training was getting worse,[14] a belief that was echoed in a study by Mehmood et al. that reviewed trainers' and trainees' understanding and perception of the curriculum changes.[15]

The Calman curriculum also differed by suggesting that numerical targets for operative experience may be imposed. These would have to be met before obtaining a CCT. The caveat that in the future it ''might be possible to assess clinical and technical competence'' appears to be the precursor to the assessments imposed in the ISCP curriculum. The Joint Committee of Surgical training has recently set new guidelines for the minimum requirement for both surgical procedures and formative assessments that a surgical trainee has to achieve before their application for the CCT.[16] Each surgical specialty has set a minimum number of surgical procedures that a surgical trainee should have performed. However, the guidelines do not specify the level of competency that a trainee has to achieve in each of the surgical procedures to be deemed competent. The same applies to formative assessments. Trainees need to complete a specified number of CBDs, CEXs, and PBAs in a variety of clinical scenarios to present a well-rounded portfolio but with the ultimate aim of demonstrating competency by the end of their training.

Addition of Work-based Assessments

Although there has been some significant progress in developing a surgical curriculum, rather than what previously appeared to be an embellished syllabus, there are still some problems. WBAs have been perhaps the most controversial, as well as universal, tools implemented within all recent medical curricula. They are known to be unpopular with trainees who feel they provide little educational benefit, and there is scant evidence to support their use as an educational tool.[17] There may be a number of factors contributing toward trainees' antipathy toward WBAs. A belief that WBAs are often inappropriately used or regarded as a ''tick-box'' exercise,[18] and that trainers do not always engage in the process and feel that they are an ineffective imposition may all contribute toward trainees' dissatisfaction with WBAs. It has been suggested that setting a minimum number of assessments was a mistake leading to the assessments being regarded as a ''mini examination'' and only being undertaken at a point when a good mark would be expected.[18] This highlights their misuse and the misinterpretation of their intention. As many of these work-based assessments are time consuming to undertake properly, trainees may feel that they want to minimize the time imposition on their trainers; similarly, trainers who are skeptical of the tool are less likely to want to carry out repeated assessments for the same procedure with the same trainee. However, it should be remembered that their value lies as a formative assessment providing feedback and allowing progression to be demonstrated.

The use of WBAs and more information regarding what should be attained with respect to knowledge and skill has made the curriculum more prescriptive. One of the drivers for change was production of a competency-based curriculum.[8] Instead, there has been documented dissatisfaction with the ISCP.[9] It is felt that the heavy emphasis on a web-based system for WBAs has been detrimental to training rather than providing validity. Competency-based training is not without problems. Placing an emphasis on this in the new curriculum has served as an argument that reduction of training time is not proving detrimental, as a minimum level of competence must be achieved. However, there has been a suggestion that placing a heavy emphasis on competency can lead to demotivation, acceptance of a minimum standard rather than striving for excellence, and increased administration costs.[19] Furthermore, focus on attaining discrete competencies, rather than taking a more holistic view to training, may compound unreadiness for independent practices.[20]

Paper vs E-portfolio

E-portfolios have started to gain increasing acceptance within medical curricula. There has been some evidence that junior doctors at the early stages of training engage well with these tools and frequently make use of them beyond the minimum requirements stipulated within their curriculum.[21] A survey of UK medical trainees demonstrated an appreciation of the importance in maintaining a record of progress.[22] The creation of an E-portfolio with WBAs marks an important difference in the 2 curricula. This was not previously used, and instead logbooks of procedures as well as trainers' reports were used to guide satisfactory progress. The emphasis on these WBAs marks a trend and shift in emphasis toward proof of competence and consequently increases the emphasis on having an accurate, convenient method of storing evidence. Having a fully integrated computerized system theoretically would encourage use and be easier to maintain allowing a PD or supervisor to access a trainees' portfolio at any time and make comments that might highlight areas of weakness and where they need to focus. In practice this does not happen, and the E-portfolio is largely reviewed only at an individuals' ARCP.

Amongst other medical programs, one important facet of the E-portfolio is as a method of recording self-reflection.

Usually this manifests as composition of reflective prose over an event or clinical experience. Within the ISCP, there is room for self-evaluation as part of the multisource feedback tool. This involves rating oneself prior to receiving confidential feedback from colleagues. The value in this may lie in providing trainers with evidence that a trainee has insight into their own abilities. Self-reflection could be developed by allowing trainees to self-assess using the same PBA or CEX tools they would ordinarily use with a trainer. This would add to a trainee's portfolio, encourage regular reflection, and help individuals identify weaknesses themselves.

CONCLUSIONS

One of the main strengths remains the syllabus component. This is a close relation to its predecessor but has been developed appropriately to include attainment levels pertinent to each individual's training level. There still remains much scope for further development. Assessment is always a contentious issue, and debate still exists regarding the validity of WBAs.[20,23] There remains dissatisfaction in how they are employed. This is one component of the curriculum that requires thought and development. As long as they are seen as an obstruction to training rather than an aid, they will continue to be poorly used. Unfortunately, often these assessments appear contrived and part of a ''tick-box'' exercise rather an educational tool. It may be that simply improving the ISCP website and its user friendliness would address much of these problems, but there also needs to be a shift in outlook from both trainees and assessors if any potential benefits from WBAs are to be acquired.

As with any curriculum, the ISCP must be continually evaluated and updated to amend any flaws and ensure that it is fit for purpose. The fundamental change in surgical education has been a shift away from a traditional apprenticeship to structured training with associated validity. On comparison with its predecessor, it has become a more compleat curriculum that seeks to address the needs of trainees and educators. However, imperfections remain and only continual evolution would allow the ISCP to meet future needs.

GLOSSARY

AES (Assigned Educational Supervisor)—consultant as trainee's base hospital responsible for progress through an academic year

ARCP (Annual Review of Competence and Progression)—formal trainee review, usually annually (although may be more frequent) to assess trainee progress. The review is undertaken by a panel including the programme director and external faculty who have access to the trainee's E-portfolio. The trainee may, or may not be present

CCT (Certificate of Completion of Training)—certificate at the end of postgraduate training to signify recognition as a specialist and that the individual may be appointed as a consultant

CEX (Clinical Evaluation Exercise)—WBA entailing an observed interaction of trainee and patient

CST (Core Surgical Trainee)—trainees on a surgical training programme (usually 2 years) undertaken after Foundation years

DOP (Direct Observation of Procedural Skills)—WBA where trainee is observed performing a minor procedure

EWTD (European Working Time Directive)—European law specifying maximum 48-hour working week

E-portfolio (Electronic Portfolio)—record of achievements including logbook based electronically—may be web based

Foundation Doctor—trainees in their first 2 years after graduation undergoing generic training within several different posts at 6-month or 4-month interval

HST (Higher Surgical Trainee)—those on defined training program leading eventually to CCT (formerly Specialist Registrars now Specialty Trainees (ST) ST3-ST8)

ISCP (Intercollegiate Surgical Curriculum Project)—current UK surgical curriculum, which is website based and allows maintenance of an E-portfolio

MMC (Modernizing Medical Careers)—UK Government Change to mechanism of speciality training, including recruitment and training programmes with greater emphasis on competency attainment

PBA (Procedure-Based Assessment)—WBA of a surgical procedure with defined components that allows evaluation of the trainee to perform the procedure as a whole, or in part

PD (Programme Director)—consultant responsible for training program for a speciality within a deanery

RITA (Record of in-training assessment)—predecessor to ARCP before 2007

WBA (Work-Based Assessment)—formative assessment with trainee assessed and given feedback by more senior colleague

REFERENCES

1. *Intercollegiate Surgical Curriculum Project*. 〈www.iscp.ac.uk〉; Accessed 5.07.2012.
2. *Calman Curriculum*. 〈http://www.jcst.org/publications/Curriculum/generalsurg_html/〉; Accessed 5.7.2012.
3. Rajan P, Din NA. Sub-specialization in general surgery—the end of the 'general' surgeon? *Br J Hosp Med*. 2005;66:185.

4. Chowdhury MM, Dagash H, Pierro A. A systematic review of the impact of volume of surgery and specialization on patient outcome. *Br J Surg*. 2007;94(2):145-161.

5. *Royal College of Surgeons*. The Surgical Specialties: 1 - General Surgery. 〈http://www.rcseng.ac.uk/media/media-background-briefings-and-statistics/the-surgical-specialties-1-general-surgery〉. Accessed 1.11.2012.

6. Thomas P. Will GP revalidation make things better or worse? *Br J Gen Pract*. 2005;55(513):318-319.

7. *General Medical Council*. Regulating Doctors, and Ensuring Good Medical Practice. 〈http://www.gmc-uk.org/education/postgraduate/merger.asp〉. Accessed 1.11.2012.

8. Jolly B, Rees L, eds. Medicel Education in the Millennium. Oxford: Oxford University Press, 1998 pp. 63-82.

9. Pereira EA, Dean BJF. British Surgeons' experiences of mandatory online work based assessment. *J R Soc Med*. 2009;102:287-289.

10. Pereira EA, Dean BJ. British surgeons' experiences of a mandatory online workplace based assessment portfolio resurveyed three years on. *J Surg Educ*. 2013;70(1):59-67.

11. Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *Br Med J*. 2010;341:c5064.

12. Lanier DC, Roland M, Burstin H, Knottnerus JA. Doctor performance and public accountability. *Lancet*. 2003;362(9393):1404-1408.

13. Bates T, Slade D. The impact of the European working time directive on operative exposure. *Ann R Coll Surg Engl*. 2007;89:452.

14. Parsons BA, Blencowe NS, Hollowood AD, Grant JR. Surgical training: the impact of changes in curriculum and experience. *J Surg Educ*. 2011;68(1):44-51.

15. Mehmood S, Anwar S, Ahmed J, Tayyab M, O'Regan D. A survey of UK surgical trainees and trainers; latest reforms well understood but perceived detrimental to surgical training. *Surgeon*. 2012;10(1):9-15.

16. *Joint Committee on Surgical Training: Guidelines for the award of a CCT in General Surgery*. 〈http://www.jcst.org/quality_assurance/Docs/cct_guidelines_gs〉; Accessed 1.11.2012.

17. Miller A, Archer J. Impact of work based assessments on doctors' education and performance a systematic review. *Br Med J*. 2010;341:c5064.

18. Beard J. Workplace-based assessment: the need for continued evaluation and refinement. *Surgeon*. 2011;9:S12-S13.

19. Leung WC. Competency based medical training: review. *Br Med J*. 2002;325:693-696.

20. Brightwell A, Grant J. Competency-based training: who benefits. *Postgrad Med J*. 2013;89:107-110. http://dx.doi.org/10.1136/postgradmedj-2012-130881.

21. Tochel C, Beggs K, Haig A, et al. Use of web based systems to support postgraduate medical education. *Postgrad Med J*. 2011;87(1034):800-806.

22. Johnson S, Cai A, Riley P, Millar LM, McConkey H, Bannister C. A survey of Core Medical Trainees' opinions on the ePortfolio record of educational activities: beneficial and cost-effective? *J R Coll Physicians Edinb*. 2012;42(1):15-20.

23. Norcini JJ, Bank LL, Duffy FD, Fortna G. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med*. 2003;138:476-481.

With all the changes that have occurred to surgical training, a comparison between the old curriculum and new curriculum was carried out to determine what the major differences might be. This review involved comparison of the curriculum that was in place, following the Calman reforms and its successor the Intercollegiate Surgical Curriculum Programme (ISCP) which was borne out of the MMC reforms. The review highlighted the fact that the new curriculum had a number of aims including:

1) Integration of the Junior Surgical Training (previous SHO grade) and the Higher Surgical Training.

2) A move towards a competency rather than time based framework.

3) The facility to better evidence learning and acquisition of skills which went hand in hand with the integral use of formative assessments.

The comparison revealed that the ISCP was significantly different from what it succeeded from the Calman reformation. Perhaps the most significant change between the two was that the ISCP strived to be an actual curriculum and incorporated not only a list of knowledge that needed to be acquired, but also details on when and how it should be obtained. Further, it included skills and a platform for recording evidence for the duration of a doctor's surgical training. This is a fundamental change from the apprenticeship model that has served trainees in the past.

Another major difference was the shift away from summative assessment- i.e. passing the FRCS coupled with spending a prerequisite period of time in training, to a competency based system where it became imperative for trainees to provide evidence in the form of "low-stakes" formative assessments.

These findings were evident from comparison and analysis of the various curriculum aspects and reviewing the rationale that led to its creation. The early parts of this chapter give a firm backdrop to how and why this evolution occurred.

There are no other evaluations of the surgical curriculum evident in published journals and none that have sought to clearly delineate what made the ISCP so different. Whilst its popularity was low at inception (and some may argue has improved little over the 10 years it has been in use) much of this was due to the scepticism that new ideas often receive, along with the emphasis on a web-based platform that was fraught with issues. Many of the teething

problems have been resolved. There are still some concerns regarding the use of WBAs, and this will be addressed in the future chapters.

2.5 Trainee Involvement in Ivor Lewis Esophagectomy Does Not Negatively Impact Outcomes- Paper and Commentary

Phillips AW, Dent B, Navidi M, Immanuel A, Griffin SM. Trainee involvement in Ivor Lewis Esophagectomy does not negatively impact outcomes. *Annals of Surgery*. 2018; 267:94-98[2]

Trainee Involvement in Ivor Lewis Esophagectomy Does Not Negatively Impact Outcomes

Alexander W. Phillips, MA, FRCSEd, Barry Dent, MA, MRCS, Maziar Navidi, MBBS, FRCS, Arul Immanuel, MD, FRCS, and S. Michael Griffin, OBE, MD, FRCSEd

Objective: The aim of the present study was to determine whether trainee involvement in esophageal cancer resection is associated with adverse patient outcomes.

Background: Operative experience for surgical trainees is under threat. A number of factors have been implicated in this leading to fewer hours for training. Esophagogastric cancer training is particularly vulnerable due to the publication of individual surgeon results and a perception that dual consultant operating improves patient outcomes. Resectional surgery is increasingly viewed as a subspeciality to be developed after completion of the normal training pathway.

Methods: Data from a prospectively maintained database of consecutive patients undergoing trans-thoracic esophagectomy for potentially curable carcinoma of the esophagus or gastroesophageal junction were reviewed. Patients were divided into 4 cohorts, according to whether a consultant or trainee was the primary surgeon in either the abdominal or thoracic phase. Outcomes including operative time, lymph node yield, blood loss, complications graded by Accordion score, and mortality were recorded.

Results: A total of 323 patients underwent esophagectomy during 4 years. The overall in-hospital mortality rate was 1.5%. At least 1 phase of the surgery was performed by a trainee in 75% of cases. There was no significant difference in baseline demographics of age, stage, neoadjuvant treatment, and histology between cohorts. There was no significant difference in blood loss ($P = 0.8$), lymph node yield ($P = 0.26$), length of stay ($P = 0.24$), mortality, and complication rate according to Accordion scores ($P = 0.21$) between cohorts. Chest operating time was a median 25 minutes shorter when performed by a consultant ($P < 0.001$).

Conclusions: These findings demonstrate that patient outcomes are not compromised by supervised trainee involvement in transthoracic esophagectomy. Training is an essential role of all surgical units and training data should be more widely reported especially in areas of high-risk surgery.

Keywords: esophagectomy, graduate medical education, morbidity, mortality, surgical outcomes, surgical training

(*Ann Surg* 2018;267:94–98)

Surgical training in the United Kingdom and the United States has moved from a traditional apprentice model to become competency based.[1,2] The European Working Time Directive has led to concerns that trainees now have fewer training hours[3–5] and this has been reflected in surgical training programs across the globe.[6,7] In addition, the challenge remains in providing both high-quality patient care and excellent training. It is known that surgical performance continues to improve after completion of training and that with fewer training hours being seen globally more importance must be placed on quality opportunities during this time.[8]

Trainee participation has been associated with both beneficial[9] and adverse,[10–12] patient outcomes. Increased autonomy has been postulated as one of the causes for adverse patient outcomes when trainees are involved.[11] Furthermore, there is an association with prolonged operating times even in those procedures that are regarded as more straightforward.[13] In contrast, there has been an association of improved colonic cancer detection when trainees are involved[9] and lower mortality rates at major teaching hospitals in patients with myocardial infarction, stroke, and heart failure.[14]

The Association of Upper Gastrointestinal Surgeons has suggested that surgeons should be carrying out a minimum of 15 to 20[15] cases per year in the United Kingdom and there has been an increasing move toward dual operating. This has the potential to reduce learning opportunities for trainees consequently necessitating that trainees carry out fellowships, as resectional surgery becomes a skill learned after completing the normal training pathway.

Esophagectomy is a technically demanding operation associated with a high level of morbidity. Several previous studies have tried to evaluate the impact of trainee involvement in esophagectomy, but have included a variety of surgical approaches with varying results.[16–18] Two of these previous studies indicated that the consultant was the main surgeon in nearly two-thirds of cases.[16,17] The aim of this study was to evaluate the impact of surgical trainee involvement in esophagectomies, looking at both short-term and medium-term outcomes.

METHOD

Patient Population

A prospectively maintained database of all patients undergoing esophagectomy for malignant disease was reviewed. Patients were treated at a single center (The Northern Oesophago-Gastric Unit, Newcastle-upon-Tyne, UK) between January 2010 and September 2013. All patients were discussed by the multidisciplinary team and subsequently underwent transthoracic (2-phase) subtotal esophagectomy, either as unimodality treatment or following neoadjuvant therapy.

Data including baseline demographics (age, sex, stage of disease, use of neoadjuvant treatment) were prospectively recorded on a standardized proforma. At the end of surgery further record of ASA grade (American Society of Anesthesiologists), blood loss, length of each phase of the operation, and who performed each phase were added to this proforma, and this is completed on discharge with details of ultimate histology, stage of disease, lymph node yield, length of stay, and complications. Complications were recorded contemporaneously and graded according to the Accordion Severity Classification of Postoperative Complications Expanded Classification system.[19] Current Union International Contre le Cancer TNM-7 was used to stage all patients (Tables 1 and 2).[20]

From the Northern Oesophago-Gastric Cancer Unit, Royal Victoria Infirmary, Newcastle-upon-Tyne, UK.

The authors declare no conflict of interests.

Reprints: Alexander W. Phillips, MA, FRCSEd, Northern Oesophago-Gastric Cancer Unit, Royal Victoria Infirmary, Queen Victoria Rd, Newcastle-upon-Tyne NE1 4LP, UK. E-mail: awphillips@doctors.net.uk.

ISSN: 0003-4932/16/26701-0094
DOI: 10.1097/SLA.0000000000002047

TABLE 1. Demographics of Each of the 4 Groups

	Group 1	Group 2	Group 3	Group 4	*P*
Number of patients	82	101	121	19	
Age	65 (21–83)	65 (30–81)	(66 (42–80)	66 (46–79)	0.43
Male	68 (84%)	85 (85%)	84 (69%)	17 (89%)	**0.017**
Histology					
Squamous cell carcinoma	17 (20.7%)	13 (12.9%)	26 (21.5%)	4 (21.1%)	0.29
Adenocarcinoma	58 (70.7%)	82 (81.1%)	283 (68.6%)	13 (68.4%)	
Other	7 (8.5%)	6 (5.9%)	12 (9.9%)	2 (10.5%)	
Neoadjuvant treatment					
None	31 (37.8%)	30 (29.7%)	44 (36.4%)	2 (10.5%)	0.30
Chemotherapy	48 (58.5%)	68 (67.3%)	70 (57.9%)	16 (84.2%)	
Chemoradiotherapy	3 (3.7%)	3 (3%)	5 (4.1%)	1 (5.3%)	
Radiotherapy	0	0	2 (1.7%)	0	
ASA					
I	13 (16%)	5 (5%)	26 (22%)	4 (21%)	**0.03**
II	43 ((53%)	61 (62%)	63 (53%)	12 (63%)	
III	25 (31%)	32 (32%)	30 (25%)	3 (16%)	
IV	0	1 (1%)	1 (1%)	0	
Pathological stage (y)p T1	5	10	11	0	0.69
T2	5	8	5	1	
T3	54	69	86	13	
T4	10	12	8	3	
Other (Tx/ HGD)	8	2	11	2	
(y)p N0	23	17	28	4	0.58
N1	27	42	39	5	
N2	27	35	41	10	
N3	5	7	12	1	

Group 1: Consultant both phases; group 2: Consultant abdomen and registrar chest; group 3: registrar abdomen and consultant chest; group 4: registrar both phases.
HGD indicates High Grade Dysplasia.
Bold values are statistically significant.

Surgical Technique

Resections were carried out using a standardized 2-phase approach (Ivor Lewis) with a radical en-bloc abdominal and mediastinal lymphadenectomy as described by Griffin et al.[21] The initial abdominal phase involved a midline laparotomy to mobilize the stomach and en-bloc lymphadenectomy. Nodal tissue along the common hepatic, proximal splenic and at the origins of the left gastric and celiac axis were resected. The lesser omentum was divided and an en-bloc resection along with hiatal dissection was performed. A gastric drainage procedure in the form of a Heineke-Mikulicz pyloroplasty was performed and feeding jejunostomy placed using a Witzel tunnel with omentopexy to the abdominal wall surrounding the jejunostomy to prevent herniation of colon into the thoracic cavity and to deal with subclinical leaks.

The thoracic phase was carried out with the patient in the left-lateral position and the right lung collapsed. The esophagus was mobilized with the middle and lower para-esophageal nodes en-bloc taking any mediastinal tissue over the pulmonary veins and aorta. Lymphadenectomy included the paratracheal, carinal, and bronchial nodes on both the left and right. The thoracic duct was resected en-bloc with the para-aortic nodes having been ligated proximally and distally. The nodes of the aortopulmonary window were also dissected out for resection. After this had been performed a ''sleeve'' resection of the lesser curve and its associated nodes was carried out—including nodes that had been dissected during the abdominal phase.

The gastric conduit was fashioned on the right gastroepiploic arcade and right gastric artery with a stapled anastomosis through the fundus at the high point of the stomach. Anastomosis was carried out using a circular stapler device (CEEA Autosuture) between (21 and 31 mm). This device was introduced through the lesser curve of the stomach, which remained open after resection. The open gastrotomy

TABLE 2. Outcomes of Each of the 4 Groups

	Group 1	Group 2	Group 3	Group 4	*P*
Number of patients	82	101	121	19	
Total operating time*	390 (210–635)	405 (270–690)	380 (220–610)	385 (295–525)	0.027
Chest operating time*	190 (120–300)	218 (140–302)	180 (110–400)	195 (150–270)	<0.0001
Abdomen operating time*	165 (85–420)	166 (91–410)	175 (90–280)	180 (90–230)	0.24
Number of resected nodes	30 (12–72)	29 (4–55)	31 (11–77)	32 (11–62)	0.26
Blood loss (mL)*	546 (465–628)	537 (470–604)	499 (439–559)	513 (377–650)	0.80
In-hospital mortality	0 (0%)	1 (0.99%)	4 (3.3%)	0 (0%)	
Complications (all)	33 (40%)	60 (59%)	56 (46%)	11 (58%)	0.21
Complications (Accordion 3+)	12 (13%)	29 (29%)	23 (19%)	3 (16%)	0.10
Length of stay*	16 (8–78)	15 (7–92)	17 (7–79)	15 (9–21)	0.24

*Median (range).
Group 1: consultant both phases; group 2: consultant abdomen and registrar chest; group 3: registrar abdomen and consultant chest; group 4: registrar both phases.

was then closed using a TA 90 stapler (Ethicon, Cinncinati, USA) and the staple line oversewn. Residual omentum from the greater curve was used to cover the anastomosis and placed between the anastomosis and trachea.

A nasogastric tube and two 24 French chest drains were placed at the end of the procedure.

Determination of Operating Surgeon

The grade of operating surgeon was recorded at the end of each operation, for each phase (abdominal and thoracic) as part of the audit proforma. In all cases surgery was performed with a consultant. A trainee was deemed to have performed a phase if they carried out greater than 75% of the procedure. In all cases this was under consultant supervision.

Analysis

For analysis purposes patients were divided into 4 cohorts according to whether a consultant or trainee performed the abdominal and chest phase.

Group 1 involved patients who had both phases performed by a consultant, in group 2 the abdominal phase was performed by the consultant and the chest by the trainee, in group 3 the abdominal phase was performed by the trainee and the chest by the consultant, and in group 4 both phases were performed by the trainee.

Complication Classification

Complications were recorded contemporaneously and entered into the database on a weekly basis. All complications were classified at the time of occurrence using the Accordion score. These were then rechecked at a monthly morbidity and mortality meeting to ensure a consensus on complication classification.

The Expanded Accordion Severity Classification of postoperative complications classifies postoperative morbidity from 1 to 6. Level 1 and 2 complications are mild and moderate complications requiring bedside or pharmacological interventions. Levels 3 and 4 indicate severe complications requiring invasive procedures without or with general anesthesia, respectively. Level 5 indicates severe organ failure and level 6 death.[19]

Statistical Analysis

Statistical calculations were performed by SPSS software, version 22.0 (SPSS, Chicago, IL).

A Kruskal-Wallis test and Mann-Whitney *U* test were used to compare continuous variables and categorical data were compared using a χ^2 Fisher exact test. A log-rank (Mantel-Cox) test was used to compare survival between groups. *P* values less than 0.05 (2-sided) were considered statistically significant.

RESULTS

Overall Results

From January 2010 to September 2013, 323 patients underwent subtotal esophagectomy via a standardized 2-phase approach. Overall mortality of this cohort was 1.5% and significant morbidity (classified as Accordion III or IV) was 20% with an overall recorded morbidity of 50%. There were 6 consultants involved in providing care and 9 trainees. Trainees were within their last 3 years of training. All had completed at least 6 years of postgraduate surgical training. The median number of esophagectomies trainees had participated in before starting at The Northern Oesophago-Gastric Unit was 8 (0–17).

The baseline demographics of those operated on was comparable between the groups. Age distribution use of neoadjuvant treatment, stage of disease with respect to pathological stage was similar. There was a significant difference between groups of sex and ASA grade. There were more females in group 3 (trainee performing the abdominal phase and consultant the chest phase) and a discrepancy in ASA between groups—with consultants more likely to perform the chest phase in patients with an ASA of 3 or above.

Consultant Versus Trainees

Trainees were the primary surgeon in part or all of 75% of cases. There was no significant difference in baseline demographics of sex, age, stage, and histology between cohorts.

There was no significant difference in blood loss ($P = 0.8$), and complication rates according to Accordion scores were not significantly different between cohorts. This was true with regards to overall complications ($P = 0.21$) and when subdivided to more severe complications (Accordion 3+) ($P = 0.1$).

Lymph node yield was high across all 4 cohorts. There was no significant difference between the groups with the range of median nodes resected from 29 to 32 nodes ($P = 0.26$). Median chest operating time was 25 minutes shorter (185 vs 210 min) when carried out by a consultant ($P < 0.001$).

There was no significant difference in length of stay across the 4 groups with median stays of between 15 and 17 days.

The only mortalities occurred in groups 2 and 3 in which the surgery was divided equally between trainee and consultant. One death occurred in group 2 and 4 deaths in group 3. In all cases the cause of death was pneumonia leading to adult respiratory distress syndrome.

Survival

Overall 2-year survival of each of the groups was reviewed. Survival was 80%, 70%, 69%, and 78%, respectively between groups 1 to 4. There was no significant difference in survival between groups ($P = 0.874$) (Fig. 1).

DISCUSSION

The findings from this study demonstrate that patient outcomes are not compromised by trainee involvement in transthoracic

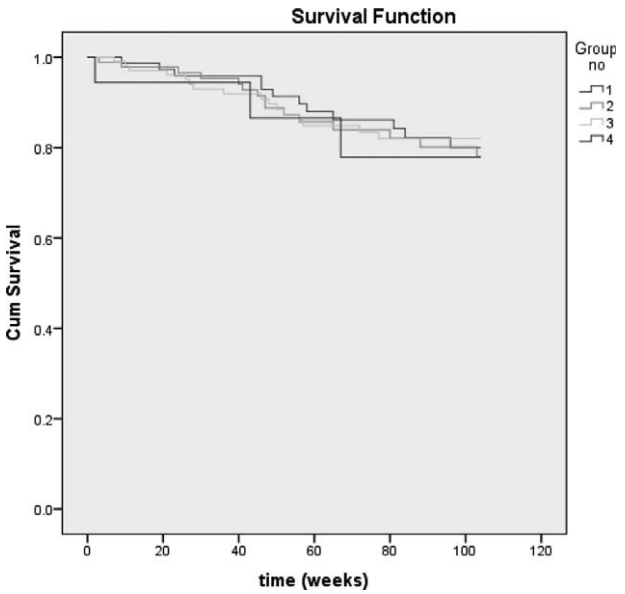

FIGURE 1. Cumulative survival of each group for the first 2 years.

esophagectomy and that trainees can perform part or all of a procedure in the majority of cases. Indeed, the only compromise that appears to exist is an extended operating time when trainees perform the thoracic phase of a 2-phase esophagectomy. As well as no significant difference in morbidity and mortality, it is also worth noting that lymph node yield and blood loss were comparable between cohorts. These are of particular note given that the need for postoperative transfusion is associated with disease recurrence,[22] and extent of lymphadenectomy implicated with long-term prognosis.[23]

A number of studies have implicated trainee involvement with adverse outcomes including increased incidence of wound infections and thromboembolic events.[24] One theory for this increase in morbidity is that it is secondary to increased operating time.[11] A further study noted that more senior trainees were likely to have increased complications and this was attributed to the ''increased autonomy'' permitted to this group.[25] This study indicates that there is no increase in morbidity with trainees operating, albeit under supervision. It should be noted that there was a median additional operating time of 25 minutes when trainees performed the chest phase in this study, but this does not seem to have clinical significance. Of note the fourth group, where the trainee completed both phases of the operation, had a lower overall operating time. This may be a factor of a smaller group potentially skewing results or a reflection of consultants consciously or subconsciously allowing trainees to complete the whole operation in patients where they suspect the procedure will be technically more straightforward. This may also account for the difference in ASA status between groups, with consultants potentially choosing to perform the chest phase in patients they deem higher risk.

Several previous studies have attempted to evaluate the impact of trainee operating on the outcomes of patients undergoing esophagectomy. Baron et al[16] reported on outcomes of 241 esophagectomies in which two-thirds were performed by a consultant. Three different operations were used in this cohort, and the only significant difference noted was a higher anastomotic leak rate in those operated by trainees. Handagala et al[17] reported a similar number of cases with a consultant as the primary surgeon and again there was no compromise to patient outcome when a trainee was the primary surgeon. Indeed, in hospital mortality was lower in the trainee group (4% vs 8%), although this did not reach statistical significance. It may be that this is due to consultants choosing to perform the entire operation in patients they deem higher risk. Putnam et al[18] evaluated outcomes of 3 techniques for esophagectomy within a resident training program. Although they did not look at specific outcomes in relation to whether a trainee performed the procedure, mortality and morbidity in this historical cohort were not deemed to be compromised by having a trainee carry out significant parts of each surgery. Furthermore, older studies by Praseedom and Paisley[26,] demonstrated that trainee involvement in complex resectional upper gastrointestinal surgery including pancreatic, hepatic, and esophagogastric resections did not negatively affect outcomes.[27]

Other surgical specialties have demonstrated that trainee involvement does not lead to poorer patient outcomes. Studies from cardiothoracic surgery have shown equivalent outcomes from trainee led lobectomies[28] and coronary artery bypass grafting.[29] Furthermore, Borowski et al[30] evaluated outcomes when trainees were involved in colorectal cancer surgery. They noted similar mortality rates between trainees and consultants with a similar case mix. There was, however, a significantly shorter hospital stay in those operated on by the trainees.

Centralization of esophageal surgery into high-volume centers has been shown to improve patient outcomes.[31] It would seem logical that this would also equate to better surgical training given the increased exposure to both operations and patient management. Fellowship programs in hospitals performing esophagectomies have been associated with reduced morbidity,[32] although it is unclear from whether the actual surgery was performed by fellows.[32] This is an interesting finding particularly given the assumption that fellows are often allowed more independence with their operating when compared with more junior trainees (residents/ registrars).[33]

The rate of complications was comparable across groups. The overall complication rate approaching 50%, the authors feel, is an accurate representation of the risks involved with esophagectomy. Data were collected contemporaneously onto a database which ensures accuracy. Individual complications were not subdivided as has been done previously. Instead use of the Accordion score aids in grouping complications into the level of impact they have on the patient. The development of a chest infection has a wide range of implications, from the need of oral antibiotics, to admission to Intensive Treatment Unit. Similarly, a chyle leak may be self-limiting and have no impact on the patient's recovery, whereas other incidences may require operative reintervention. These data suggest that complications that were graded as more ''severe'' were again comparable between groups.

It is also worth noting that there is no significant difference in survival between patients over the first 2 years. This is an important consideration, as a number of issues have been associated with survival, from lymph node yield to complications. Two- year survival without stratification into pathological stages is 70% to 80% between groups. This represents a high-potential cure rate, and suggests there is no compromise when a trainee performs part of the operation.

The major strength of this study is that it evaluates outcomes for a single operation performed in a standardized fashion at a high-volume unit. All operations were supervised by a consultant whether or not they were the primary surgeon for that phase of the operation. High-volume units allow not only for improved surgical outcomes, but also provide an opportunity for trainees to hone and develop both surgical skills, and their perioperative management. Markar et al[34] recently suggested that a threshold of 15 cases existed for consultants ,which reflected an improvement in 30-day mortality and that 1-, 3-, and 5-year outcomes were affected by volumes of 35 to 59 cases. This study, however, did not factor in the training that surgeons had received before taking up independent practice.[35] Indeed improvement in outcomes has been demonstrated before,[8] but the impact of a consultant learning curve can be ameliorated by increasing the experience and exposure while still in training. Training at high-volume centers, we believe, has a profound impact on future practice and is likely to influence the management of complicated patients. A wealth of experience obtained in training is likely to shift any learning curve to the left, but needs to be accompanied by a high-volume practice to maintain skills.

One limitation of this study is the disparity in group sizes with only 6% of operations being completed entirely by a trainee. Having unequal groups may affect the power to identify differences when they actually exist. It should also be noted that all procedures were carried out with a consultant and trainee and so it is not possible to comment on whether 2-consultant operating would alter outcomes.

The experience of trainees in esophageal surgery before starting work at the unit was limited. This was highlighted by the low median number of cases (8) before starting at the unit. The purpose of this study was, however, not to establish learning curves but to demonstrate that training in this complex surgery is possible within the confines of a surgical training program.

Another important consideration is informed patient consent. All patients have to know that a trainee will perform part, if not all, of the procedure. Although patients seem generally supportive of trainee involvement, there has been a suggestion that as they

are made more explicitly aware of their involvement they are less likely they are to give consent.[36] Similarly there is evidence that patients are more likely to consent to participation of more senior trainees than junior trainees.[36] Most patients want to be informed of the risks involved in surgery and it is important to be able to discern whether any additional risk exists with trainee participation.

This study shows that there is no detrimental impact to trainees carrying out significant components of an esophagectomy while supervised. Excellent results can be obtained with no additional morbidity or mortality and there appears to be no compromise in oncological outcome when viewed either by lymph node yield or by 2-year survival. It should serve as a standard for resectional units and be an additional marker in national audits of excellence.

REFERENCES

1. Phillips AW, Madhavan A. A critical evaluation of the intercollegiate surgical curriculum and comparison with its predecessor the ''Calman'' curriculum. *J Surg Educ*. 2013;70:557–562.
2. Sonnadara RR, Mui C, McQueen S, et al. Reflections on competency-based education and training for surgical residents. *J Surg Educ*. 2014;71:151–158.
3. Cowie CJA, Pešić-Smith JD, Boukas A, et al. Has the impact of the working time regulations changed neurosurgical trainees' attitudes towards the European Working Time Directive 5 years on? *Br J Neurosurg*. 2013;27:580–585.
4. Maisonneuve JJ, Lambert TW, Goldacre MJ. UK doctors' views on the implementation of the European Working Time Directive as applied to medical practice: a quantitative analysis. *BMJ Open*. 2014;4:e004391.
5. Parsons BA, Blencowe NS, Hollowood AD, et al. Surgical training: the impact of changes in curriculum and experience. *J Surg Educ*. 2011;68:44–51.
6. Gough IR. The impact of reduced working hours on surgical training in Australia and New Zealand. *Surgeon*. 2011;9(suppl 1):S8–S9.
7. Antiel RM, Reed DA, Van Arendonk KJ, et al. Effects of duty hour restrictions on core competencies, education, quality of life, and burnout among general surgery interns. *JAMA Surg*. 2013;148:448–455.
8. Sutton DN, Wayman J, Griffin SM. Learning curve for oesophageal cancer surgery. *Br J Surg*. 1998;85:1399–1402.
9. Buchner AM, Shahid MW, Heckman MG, et al. Trainee participation is associated with increased small adenoma detection. *Gastrointest Endosc*. 2011;73:1223–1231.
10. Iannuzzi JC, Chandra A, Rickles AS, et al. Resident involvement is associated with worse outcomes after major lower extremity amputation. *J Vasc Surg*. 2013;58:827–831.
11. Kasotakis G, Lakha A, Sarkar B, et al. Trainee participation is associated with adverse outcomes in emergency general surgery: an analysis of the National Surgical Quality Improvement Program database. *Ann Surg*. 2014;260:483–490.
12. Aguilar B, Sheikh F, Pockaj B, et al. The effect of junior residents on surgical quality: a study of surgical outcomes in breast surgery. *Am J Surg*. 2011;202:654–657.
13. Papandria D, Rhee D, Ortega G, et al. Assessing trainee impact on operative time for common general surgical procedures in ACS-NSQIP. *J Surg Educ*. 2012;69:149–155.
14. Polanczyk CA, Lane A, Coburn M, et al. Hospital outcomes in major teaching, minor teaching, and nonteaching hospitals in New York state. *Am J Med*. 2002;112:255–261.
15. Association of Upper GI Surgeons of Great Britain and Ireland. Guidance on minimum surgeon volumes. 2010. Available at: http://www.augis.org/wp-content/uploads/2014/05/AUGIS_recommendations_on_Minimum_Volumes.pdf. Accessed April 5, 2016
16. Baron R, Sujendran V, Maynard N. Should oesophagectomies be performed by trainees? The experience from a single teaching centre under the supervision of one surgeon. *Ann R Coll Surg Engl*. 2008;90:305–309.
17. Handagala SDM, Addae-Boateng E, Beggs D, et al. Early outcomes of surgery for oesophageal cancer in a thoracic regional unit. Can we maintain training without compromising results? *Eur J Cardiothorac Surg*. 2012;41:31–34.
18. Putnam JB, Suell DM, McMurtrey MJ, et al. Comparison of three techniques of esophagectomy within a residency training program. *Ann Thorac Surg*. 1994;57:319–325.
19. Strasberg SM, Linehan DC, Hawkins WG. The accordion severity grading system of surgical complications. *Ann Surg*. 2009;250:177–186.
20. Sobin LH. *TNM Classification of Malignant Tumours.* 7th ed. Hoboken, New Jersey: Wiley-Blackwell; 2009.
21. Griffin SM, Raimes SA, Shenfine J, eds. *Oesophagogastric Surgery: A Companion to Specialist Surgical Practice*. 5th ed., Philadelphia: Elsevier; 2013:94–105. Chapter 5.
22. Dresner SM, Lamb PJ, Shenfine J, et al. Prognostic significance of perioperative blood transfusion following radical resection for oesophageal carcinoma. *Eur J Surg Oncol*. 2000;26:492–497.
23. Phillips AW, Lagarde SM, Navidi M, et al. Impact of extent of lymphadenectomy on survival, post neoadjuvant chemotherapy and transthoracic esophagectomy. *Ann Surg*. 2016. In Press. doi: 10.1097/SLA.0000000000001737.
24. Krell RW, Birkmeyer NJO, Reames BN, et al. Effects of resident involvement on complication rates after laparoscopic gastric bypass. *J Am Coll Surg*. 2014;218:253–260.
25. Scarborough JE, Bennett KM, Pappas TN. Defining the impact of resident participation on outcomes after appendectomy. *Ann Surg*. 2012;255:577–582.
26. Praseedom RK, Paisley A, Madhavan KK, et al. Supervised surgical trainees can perform pancreatic resections safely. *J R Coll Surg Edinb*. 1999;44:16–18.
27. Paisley AM, Madhavan KK, Paterson-Brown S, et al. Role of the surgical trainee in upper gastrointestinal resectional surgery. *Ann R Coll Surg Engl*. 1999;81:40–45.
28. Chaudhuri N, Grayson AD, Grainger R, et al. Effect of training on patient outcomes following lobectomy. *Thorax*. 2006;61:327–330.
29. Goodwin AT, Birdi I, Ramesh TP, et al. Effect of surgical training on outcome and hospital costs in coronary surgery. *Heart*. 2001;85:454–457.
30. Borowski DW, Ratcliffe AA, Bharathan B, et al. Involvement of surgical trainees in surgery for colorectal cancer and their effect on outcome. *Colorectal Dis*. 2008;10:837–845.
31. Birkmeyer JD, Siewers AE, Finlayson EVA, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128–1137.
32. Kohn GP, Galanko JA, Meyers MO, et al. National trends in esophageal surgery—are outcomes as good as we believe? *J Gastrointest Surg*. 2009;13:1900–1910.
33. Rohatgi A, Sutcliffe R, Forshaw MJ, et al. Training in oesophageal surgery—the gold standard: a prospective study. *Int J Surg*. 2008;6:230–233.
34. Markar SR, Mackenzie H, Lagergren P, et al. Surgical proficiency gain and survival after esophagectomy for cancer. *J Clin Oncol*. 2016;34:1528–1536.
35. Phillips AW, Dent B, Navidi M, et al. No T. *J Clin Oncol*. 2016; doi: 10.1200/JCO.2016.68.1817. Epub ahead of print August 29.
36. Porta CR, Sebesta JA, Brown TA, et al. Training surgeons and the informed consent process: routine disclosure of trainee participation and its effect on patient willingness and consent rates. *Arch Surg*. 2012;147:57–62.

Hitherto, the evolution of the surgical curriculum has been addressed. There have been significant changes to surgical training throughout the globe, generally resulting in fewer hours to learn the necessary skills, and also fewer opportunities to receive training. This manuscript aimed to determine whether these changes have been detrimental to trainees in the UK.

Oesophagectomy is a technically demanding procedure associated with significant morbidity. In the last national audit there was an associated mortality of 3% and large volume units report complication rates at approximately 50%. There have been concerns that the learning of more complicated skills has moved outside the remit of a normal training programme and that trainees that wished to carry out such surgeries would necessarily need to participate in a specialist fellowship year.

Another important consideration is the impact of trainees operating on patients on surgeon outcomes. The publication of surgeon outcomes may conceivably compromise patient care as well as training. Patient care may be compromised as surgeons potentially stay away from performing higher risk surgery due to the risk it will affect outcomes and be highlighted in national audits. Training may be compromised because trainers are concerned that allowing a trainee to perform all, or part of an operation, may adversely affect the overall outcome.

This study involved data from a unit with a high volume of oesophageal resections. Data was collected in a prospective fashion ensuring the recording of complications and other outcome parameters were accurate. It demonstrated that a standardised oesophagectomy could be performed by trainees safely. There was no detrimental outcome by allowing a trainee to perform a single phase or all of the operation. Indeed the only significant difference was that trainees were slower at performing the thoracotomy phase of this surgery.

Whilst this study demonstrates that such a complex procedure can be taught within the confines of a normal training programme, it does not look at what the original intention was at the start of the procedure. Whilst the unit involved has a strong ethos of training, no documentation was made of why consultant or trainee performed the phase of the operation that they did. Thus an unconscious bias is likely to exist where a consultant perceives a particular phase is likely to be more difficult, or indeed the whole case more difficult, and chooses to do that section themselves. Similarly there is no documentation of when trainees may have commenced a case but due to failure to progress the consultant took over as the main surgeon. This study took a pragmatic approach and aimed to demonstrate safety in

allowing trainees to carry out complex cases under consultant supervision. Whether there was an unconscious or deliberate decision by a consultant to do a particular part, or all of the surgery, the impact was excellent outcomes. Only 6% of procedures were carried out completely by a supervised trainee. It is difficult to give a definite reason why this was so low. However, the length of the procedure, with two natural phases lends itself to a change in primary operator, and the consultants present may feel that they should carry out one of these phases. There may also be a perception that trainees take longer to carry out a stage. This was borne out by the findings that they do take significantly longer to carry out the chest phase. It is imperative that sufficient time is given to trainees to allow them to learn, provided this does not compromise patient outcomes. Splitting the case between trainee and trainer does potentially allow for this.

The findings of this paper have implications for patient consent. Patients are aware that "trainees" are involved in their procedures, but evidence that supervised trainees fare no worse than the consultant performing the procedure may significantly alleviate any worries a patient may have.

However, possibly more striking is the impact this manuscript has in supporting trainers and ensuring that they continue to allow trainees to operate, particularly in more complicated cases. Whilst the outcome of this paper was not to evidence that trainees were competent to independently perform the procedure at the end of a placement within this unit, nor was it to try and establish a learning curve, what it reiterated was that training is an essential role of any surgical unit. It also suggests that training data should be incorporated into national audits. This will highlight the units that are striving to provide training, an important issue when some institutions are moving to dual consultant operating for complex cases.

2.6 Chapter Conclusion

As a result of a number of factors, pressures on time to train, a drive to streamline training that was in existence, and a desire to produce surgeons with evidence of competencies, surgical training has changed vastly over the last 10 years.

What now exists in the UK is a more complete curriculum, which is held together using an online platform that explicitly stipulates trainee learning outcomes. In addition, it identifies when these should ideally be achieved by surgical trainees.

Despite all the concerns regarding surgical training in the UK it is possible to adequately achieve a high level of operating ability within a training programme. Extra time spent in a fellowship, honing skills, has become increasingly popular. This should not be used to reduce the emphasis on training in more advanced skills within the specialty training programme, which can be achieved without compromise to patient care.

.

Chapter 3: Performance Assessment and Feedback of Technical Skills and Simulation

Theme 2: Performance assessment and feedback of technical skills

1) Phillips AW, Jones AE. Workplace based assessments in surgical training: Valid and reliable? *Bulletin of The Royal College of Surgeons of England* 03/2015; 97(3):e19-e23[3]
2) Nesbitt CI, Phillips AW, Searle RF, Stansby G. Randomised trial to assess the impact of supervised and unsupervised video feedback on teaching practical skills. *J Surg Educ*. 2015 Jul-Aug;72(4):697-703[4]
3) Phillips AW, Bookless LR, Matthan J, Whitehead I, Madhavan A, Nesbitt CI, Stansby G. Individualised expert feedback is not essential for improving basic clinical skills performance in novice learners: A Randomised Trial. *J Surg Ed* 2017 Jul-Aug; 74(4):612-620[5]

This chapter of the thesis looks at two aspects of assessment. The use of WBAs, which are an integral part of both foundation programmes for junior doctors, and also part of the ISCP is included and the first paper reviews the validity and reliability of WBAs.

The second part of this chapter looks at the use of simulation in learning technical skills. A background to simulation is provided and the use of video feedback in a "novice" group of medical students from two randomised trials constitutes the two final papers.

3.1 Performance Assessments within Surgical Training.

3.1.1 Assessment within Surgical Training

Use of WBAs has become increasingly popular within the surgical curriculum. They are used as a formative, rather than summative assessment and help to evaluate the progress of a trainee. Usually a trainee will meet with an educational supervisor at the start of any attachment. At this meeting training objectives are set using the framework of the intercollegiate surgical curriculum project (ISCP). As part of the ISCP various WBAs including mini Clinical Evaluation Exercises (mini-CEXs), Case Based Discussions (CBDS), directly observed procedural skills (DOPS) and procedural based assessments (PBAs) are expected to be completed in order to aid the trainee in demonstrating progression and achievement of competencies.

One of the drivers for the use of WBAs is the aim set by the Post Graduate Medical Education Training Board (PMETB) to produce a competency based curriculum[55]. However, in striving to achieve a pre-set list of competencies there is a risk that doctors can end up focusing on attaining a set of WBAs as evidence. This may detract from trainees actually becoming competent.

At their inception, these assessments were often regarded with suspicion and as a high stakes event, rather than a learning opportunity. They have still not been fully accepted by trainees and trainers alike. Thus the question of how much benefit they confer, and whether they are being used properly remains. Most regional deaneries stipulate a minimum number of each type of assessment that must be completed over the training year. This potentially places an emphasis on quantity, rather than quality learning events.

The importance of WBAs is immediately apparent. They are used at annual reviews of trainees as evidence and to decide if career progression can occur. It is therefore important to ensure their validity as a formative assessment. Further, the actual process of carrying out a WBA, if carried out properly, may be time consuming for both trainer and trainee. Thus if these tools are to cement their place within the surgical curriculum it is also important to establish their effectiveness.

3.1.2 Defining assessment

3.1.2.1 Formative Assessments

Formative assessments are designed not to provide a distinct mark but instead aid trainers in providing feedback on progress. Their aim is to support the curriculum they have been integrated into, enhance the learning, and allow constructive feedback for the student. This last issue is of vital importance in learning and a feature that is often lacking in summative assessments[56] and at the "heart of medical education"[57].

Feedback has been postulated as helping learning in three ways[58,59].

1) Informs students of their progress.

2) Advises students of learning needs, and resources available to facilitate learning.

3) Motivates students to participate in appropriate learning activities.

However, there has been suggestion that there may not necessarily be a correlation between formative and summative assessments. Anziani et al found no correlation in a cohort of undergraduates in oral surgery examined in both practical skills and history taking and examination[60]. They speculated on several reasons for this including 1) different levels of anxiety between the summative assessment and the formative assessment; 2) the summative assessment evaluated the whole procedure using an objective list whilst the formative episodes did not; 3) the summative assessment looked at an end point whilst the formative assessment monitors progress from novice to competence.

3.1.2.2 Summative Assessment

Summative assessments are regarded as a high stakes terminal evaluation. Within the surgical curriculum this occurs at two points. The MRCS exam is attempted in the junior years, which must be passed as a prerequisite for entry into higher surgical training. The FRCS exam constitutes the final summative exam that is taken towards the end of surgical training and exists as a final assessment that must be passed in order to complete surgical training.

Whilst a combination of formats is used throughout these exams: Multiple choice questions (MCQ), Extended matching questions (EMQ), Objective Structured Clinical Examinations

(OSCEs) and viva and clinical exams, the end aim is to determine learning. These high stakes exams are hugely motivating for students given the importance placed on them before progress can be made.

3.1.3 Value of Feedback

Provision of feedback to trainees can be difficult, and often may not be provided in an optimum fashion during the course of the normal working day. There may be several reasons preventing effective feedback. These include lack of time, or an opportune moment; reluctance on the part of the teacher to cause offence or a defensive response; or simply ineffective feedback. Whilst WBAs could be thought of as a contrived and artificial situation the cases are real patients and in the example of CBDs value can be derived from working through the diagnosis and management plans.

Feedback has been demonstrated to be beneficial in the development of any trainee. Feedback has been shown to change trainees' behaviour. Hattie and Timperley summarised information from 12 meta-analyses suggesting feedback had an effect size of 0.79 (increased the mean on an achievement test by 0.79 of a standard deviation)[61]. Veloski et al carried out a systematic review of 41 studies with 74% of these demonstrating a positive effect for feedback alone[62].

Students' behaviour has also been shown to be affected by the provision of feedback and carrying out WBAs. Burch et al found that those medical students carrying our frequent DOPs were more likely to carry out encounters without feeling they needed to consult the notes in advance while traditionally students would interview patients only after consulting notes[63]. The use of WBAs was also reported to increase the likelihood of students reading around a subject and revising the topics that arose in the assessment[63].

3.1.4 Validity and Reliability of Assessment. Paper and Commentary.

Phillips AW, Jones AE. The validity and reliability of workplace-based assessments in surgical training. *Bulletin of The Royal College of Surgeons of England* 03/2015; 97(3):e19-e23[3]

The validity and reliability of workplace-based assessments in surgical training

Will we ever see WBAs as more than a box-ticking exercise?

AW Phillips Registrar in General Surgery[1]
AE Jones Registrar in General Surgery[2]
[1]Newcastle upon Tyne Hospitals NHS Foundation Trust
[2]University Hospital Southampton NHS Foundation Trust

DOI: 10.1308/147363515X14134529301147

The use of workplace-based assessments (WBAs) as a method of assessing doctors' competence has increased in popularity throughout all postgraduate medical specialties during the past decade.[1] The need for objective measurements of doctors' competence and performance has arisen as a result of a number of issues. These include a perceived loss of trust between the general public and medical professionals, and the new system of 'fast tracking' foundation doctors into specialty training instituted through Modernising Medical Careers.[1] Fully implemented in the UK in 2009, the European Working Time Directive (EWTD) has led to an increase in trainees' work intensity but a fall in trainees' quality learning opportunities, particularly in surgical subspecialties.[2,3] This, in turn, has led to reduced opportunities for personal reflection and feedback from colleagues.[4]

All have contributed to a large and widespread change in medical training and assessment during the past few years. The impact of these changes has been felt across all medical disciplines and subspecialties, with pressure mounting on the medical specialty as a whole to provide objective evidence of trainee competence and performance.[5,6] WBAs were introduced to address this issue.

WBAs reflect a more objective measure of trainee competence. They move away from competence presumed to have been achieved simply by reaching a particular level of seniority or from working for a particular time in that subspecialty. In addition, it is argued that WBAs are directly able to measure trainee performance (ie what the trainees actually do 'in real life'). In this way, they have been said to complement more traditional 'cognitive' assessment methods such as the written and *viva voce* examinations imposed by the royal colleges.

A further benefit of WBAs includes their recent integration into a General Medical Council (GMC)-approved curriculum, following the merging of the Postgraduate Medical Education and Training Board (PMETB) with the GMC in 2010. As well as facilitating a consistent approach to training across all subspecialties, it was hoped that this would provide evidence towards revalidation and also highlight underperforming doctors who may need further training.

Competence and performance

Competence and performance differ: Competence requires appropriate knowledge, skills and attitudes, all of which are required in some form to be a competent surgeon. Assessments that measure competence have a specific set of criteria and outcomes against which a trainee is compared. Tests of competence assess at the penultimate level ('shows how') of Miller's pyramid.[7] Consequently, trainees may be regarded as competent if they satisfy a number of predetermined criteria (a 'checklist'), usually in a simulated scenario or during a discussion with a trainer.

In contrast, assessing performance is assessing competence to perform the same tasks as previously but this time in the workplace and under the stresses of a normal working environment. It measures what a trainee actually does, in real life, and corresponds to the highest tier of Miller's pyramid ('does'). It is the closest form of assessment to actual practice.

The need for objective assessments of both competence and performance have led to a wide range of WBAs. These attempt to assess a wide range of skills and attributes required to be a safe and independent practitioner in that subspecialty.

FORMATIVE AND SUMMATIVE ASSESSMENTS

In principle, there are two forms of assessment. Formative assessments (*for* learning) are continuous, informal, 'low-stakes' assessments that should ideally be performed regularly with numerous assessors. They allow for continuous feedback and improvement, and the more that are completed, the better, with reliability correlating well with number performed. Formative assessments should therefore be performed on more than one occasion, should be assessed by more than one trainer and, where possible, by more than one formative assessment tool (the process of 'triangulation').

In contrast, summative assessments (*of* learning) are formal, infrequent, 'high-stakes' assessments such as examinations, when the result is pass or fail. Written exams and *vivas*, the mainstay of the royal college examinations, traditionally test more cognitive processes (comprising the lower two tiers of Millers pyramid – 'knows' and 'knows how'). WBAs complement these more traditional testing methods by regularly assessing the application of this knowledge and the competence in performing the task in a workplace environment (comprising the upper two tiers of pyramid – 'shows how' and 'does').

A number of WBAs have been integrated into the Intercollegiate Surgical Curriculum Programme as formative assessments to aid the provision of feedback. However, there is increasing evidence to suggest that they are frequently used incorrectly, as summative rather than formative assessments, to evaluate progress of a trainee.

Part of the reason for the use of WBAs is the requirement of the PMETB to produce a competency-based curriculum.[8] Nevertheless, there is arguably a danger that doctors can end up striving to achieve a list of competencies, which may potentially detract from striving for excellence.

There are four main types of WBA used in the surgical curriculum. Each is designed to help provide formative feedback on different aspects of surgical skills. The direct observation of procedural skills (DOPS) is designed to promote development of practical skills. Procedures are identified as appropriate skills to have developed at each stage of training and the trainee will perform these while being observed by a trainer. As the trainee advances, the DOPS is ultimately superseded by the procedure-based assessment (PBA), which assesses the ability of the trainee to perform all or part of a surgical procedure.

Case-based discussions (CBDs) involve trainees using real patients and notes as a focus of discussion. The idea is that the trainer can explore the trainees' ability to apply their knowledge when carrying out appropriate clinical reasoning and decision making. It may not necessarily focus on clinical themes, allowing professional and ethical aspects to be explored.

Finally, the mini-clinical evaluation exercise (mini-CEX) involves the direct observation and assessment of the trainee during a clinical encounter, the focus of which may be history taking, clinical examination, communication or overall clinical judgement. It is designed to provide and allow feedback on a 'snapshot' of a typical trainee–patient interaction. Again, as with DOPS, these were originally intended as an assessment tool for junior trainees but are now also required for higher surgical trainees.

WHAT MAKES A GOOD ASSESSMENT TOOL?

The ideal assessment tool should satisfy a number of criteria. These were described originally by Van Der Vleuten and formulated into a 'utility index', a mechanism used later for assessment design and evaluation.[9]

Validity and reliability are the two key components.[10] Other important factors include cost effectiveness, acceptability and educational impact. Feasibility is an additional attribute added subsequently although it can be regarded as being implicit in cost effectiveness and acceptability.

Validity describes the ability of a test to measure the attribute it sets out to measure. As a result, multiple choice questions are a valid method of assessing knowledge as they are able to cover a wide range of topics. Several different types of validity exist, which include face validity, content validity, construct validity and criterion validity.

Face validity looks at how closely the simulator resembles the real world situation and is largely an intuitive judgement. Content validity, on the other hand, relies on looking at how closely the scenario reflects what it sets out to measure (ie the particular skill involved). Construct validity refers to whether the scenario correlates to a theoretical model. Outcomes should, in theory, be related such that a more experienced surgeon would be expected to do better than a novice at the skill involved. Criterion validity can be subdivided into predictive validity (how well ability in the scenario predicts future performance) and concurrent validity (the

correlation between the assessment tool and the 'gold standard').[10]

Reliability describes the ability of a test to achieve consistent, reproducible results again and again. Achieving a result 'guarantees' a particular standard of competence. Theoretically, students of the same level should perform equally well. Similarly, the task should present the same level of challenge on each attempt. That is not to say that the scenario will not appear easier as it is practised more. The number of procedures observed for a test to be regarded as reliable is controversial but there is evidence that a trainee should be observed by at least three trainers (each of whom should observe two procedures) to allow adequate reliability.

Thus, the ideal assessment tool must actually test the skillset that it was originally designed to test, and also be consistent and reproducible.

However, despite good evidence for their reliability in formative assessment, WBAs are not regarded as being reliable enough to be used as a 'standalone' assessment of competence.[11] Current evidence suggests that they should be used along with more traditional, summative endpoint assessments, in the form of examinations. In this way, the upper two tiers of Miller's competence pyramid ('shows how' and 'does') are covered.

CBDS

The CBD evolved from the American Board of Emergency Medicine's chart-stimulated recall (CSR).[12,13] It has been included as a WBA in the UK foundation programme as well as for several medical specialties. CBDs involve a structured discussion between trainer and trainee about a recent case, perhaps familiar to both parties or to whom the trainee has had the responsibility of attending. They are designed to assess clinical judgement, decision making and the application of surgical knowledge.

The CSR assessment was found to be a powerful tool for assessing performance.[14,15] A modified version, the case-based oral assessment, was implemented by the GMC to assess seriously deficient doctors. Being a generic instrument, it could be adapted to each specialty as required.[16] Those being reviewed needed to supply a portfolio of cases to two assessors. Those piloting the system felt the tool allowed for a good representation of actual daily activity. However, this British version of the CSR assessment contrasted with its US counterpart as it was a summative tool for doctors in a stressful (failing) situation, using two assessors (as opposed to one).

The case-based oral assessment evolved into the CBD. The intention of the CBD was to allow trainees to demonstrate knowledge, reasoning and decision making. Patient records are used as the basis of discussion, allowing clinical record keeping to also be assessed. Paediatric trainees studied by Mehta *et al* felt that CBDs were a valuable tool, providing an opportunity for quality learning and feedback,[17] with similar conclusions being drawn from a study of CBD use in oncology training[18] and general practice training.[19] Nevertheless, the latter study in general practice trainees concluded pertinently that CBDs are of limited use as an assessment in isolation. On the other hand, good feedback did aid the trainees in improving their performance in future assessments, although this was dependent on the skill and confidence of the educator.

The logbook is now known to be an unreliable tool

Across specialties, it appears that CBDs are well regarded by trainees, as long as the trainer engages with them and provides appropriate feedback. There has been little research into how outcomes from these assessments compare with actual clinical performance or other WBAs. They do, however, have face validity and allow trainers to explore higher thinking.

DOPS

DOPS is the observation of a technical or surgical procedure (or part thereof) on a patient. As the trainee carries out the procedure, the assessor's evaluation is recorded on a structured form, a checklist of defined essential steps. Trainers can provide immediate feedback on the trainee and tailor their subsequent teaching practice towards areas in which the DOPS performance suggests the trainee requires more guidance.

The DOPS assessment has now largely superseded the objective structured clinical examination (OSCE), introduced originally for assessing practical skills competence. In the past, the OSCE has been criticised for assessing procedures in a fragmented manner rather than assessing the complete procedure, usually owing to time constraints in an artificial examination scenario.

A related assessment tool, the objective structured assessment of technical skills, was developed in Canada and the US in 1995 for the assessment of surgical residents. It assesses trainees' competence in performing six general surgical procedures, including excision of skin lesion and bowel anastomosis.[20]

DOPS is especially relevant to surgical subspecialties, where trainees can ultimately be expected to be able to perform a wide range of practical procedures. This repertoire will include both simple and complex surgery, as well as elective and emergency surgery deemed traditionally to be covered in the surgical logbook. Conversely, the logbook by itself is now known to be an unreliable tool in the assessment of procedural competence.

DOPS is regarded as being a valid assessment tool as it directly examines the upper tier of Miller's pyramid ('shows how'). In particular, it satisfies the face validity criteria,

rating trainees in their workplace performing real tasks on real patients.

However, although DOPS may be perceived by trainees and assessors as being more reliable and valid than previously existing assessment instruments, there is very little evidence for this in the literature. In addition, DOPS potentially poses significant time constraints as the time taken to complete a DOPS assessment generally equals the time taken to perform the actual procedure, as well as the subsequent time taken to provide and document feedback. For simpler procedures, mainly relevant to junior trainees (eg cannulation), this may not be an issue. On the other hand, for more senior trainees performing more complex procedures such as colonoscopies, they may find it increasingly difficult and frustrating to achieve adequate numbers of DOPS assessments as specified by deaneries in preparation for their annual review.

PBAs

PBAs were adapted originally from tools used to assess orthopaedic trainees' surgical performance.[21] Surgical PBAs therefore aim to assess technical, operative and professional skills within a wide range of elective and emergency surgery. They are designed for more technically involved procedures (usually actual operations that are divided into many steps).

This assessment process begins typically with consenting a patient, moves through liaising with theatre staff and preparing a patient, to actually performing the procedure (which itself may be divided into several steps) and constructing an operation note with postoperative instructions. Each part of the task can be graded as N for 'not applicable' or 'not observed', S for 'satisfactory' (ie at the level required for the completion of training) and D for 'development required'. As the PBA potentially assesses multiple parts of an operation, it can be used to focus and provide feedback on a number of different aspects of the procedure.

MINI-CEX

The mini-CEX is designed to assess clinical skills during actual encounters with patients. It requires multiple assessments to be reliable. However, over the years, it has been used in a wide range of contexts and compared with other assessments across programme years, making it difficult to establish its validity.[22] Despite this, the mini-CEX does have good construct and criterion validity based on comparisons between trainees' achievement in these assessments and clinical skills performance.

There have been no studies of the use of the mini-CEX in a surgical cohort. Brazil *et al* looked at the cost effectiveness of implementing the use of the mini-CEX in an emergency medicine setting.[23] They felt that the tool was time consuming and therefore costly, and that it led to a decline in the department's efficiency. It also produced no additional benefit to the current in-training assessment in identifying underperformance. Nevertheless, both the assessors and the trainees felt the mini-CEX provided a valid assessment of performance.

the face-to-face nature of the feedback may encourage lenient scoring

In contrast, in a study looking at the use of the mini-CEX in anaesthetic trainees, trainers and trainees found that the assessment was easy to incorporate into the working day, and perceived it to have a positive educational impact.[24] While both these groups felt that the mini-CEX was a useful tool, a study looking at UK foundation doctors revealed that they felt the mini-CEX was not a useful component of their training, which might reflect that only a minority had training in using the tool.[25] Another reason given for the negative responses towards the mini-CEX was a lack of understanding by trainers and trainees about the formative purpose of the tool, leading them to use it as a tick-box exercise rather than a learning event.

Durning *et al* looked at the validity of the mini-CEX in medical training by comparing results with two established WBAs.[26] They felt that it has established validity owing to its correlation with other tools.

As with all formative assessments, the value in the mini-CEX lies in the quality of the feedback given. After looking at feedback in general practice and trainee assessment pairs, one study suggested that the feedback is determinant on the users rather than the tool.[27] They further suggested that in order to enhance the effectiveness of these assessment tools, there needs to be an emphasis on reflection and producing action plans. A further study looking at the use of the mini-CEX in anaesthetic training found that reliability was limited by assessor variability.[24] There are a number of reasons postulated for lack of discriminatory feedback. Trainers may adjust the feedback to the level of training of the trainee rather than a predefined level (usually a level of competence), particularly showing lenience towards junior trainees in difficult cases.[26] Similarly, the face-to-face nature of the feedback may encourage lenient scoring.[24]

WBAs: SHORTFALLS AND IMPROVEMENTS

WBAs were planned as formative assessment tools, which benefited trainees by providing continuous feedback during the course of training. They were intended to complement the formal summative examinations. Along with other assessment tools such as multisource feedback and reflective e-portfolios, there is good evidence that they can be enabling for trainees, helping them to achieve educational objectives.

However, as trainers and trainees alike become familiar with these newer forms of assessment, there is evidence that WBAs are not being used in the manner in which

they were intended, thereby limiting their educational value. There is also evidence of a lack of engagement by trainers, and a perceived lack of validity and reliability by both trainers and trainees.[28,29]

The WBAs are often regarded as too time consuming in an era with increasing service commitments, for trainees as well as the trainer. Paradoxically, their reliability in assessing a trainee increases with an increasing number completed. With EWTD-compliant rotas restricting surgical training and some deaneries now requiring up to 80 WBAs per year, there is a suggestion that WBAs are being misused. Furthermore, there is concern that they are being used incorrectly in a summative rather than formative manner.

In addition, instead of being performed continuously, there is a tendency for trainees to conduct WBAs later in their attachment, when they are more accomplished and more likely to score highly. This makes it difficult to demonstrate a trainee's progress during that attachment. In some instances, there are reports of WBAs being filled in *en masse* during a single meeting so as to achieve stipulated deanery targets, thereby defeating the entire foundation on which they are based.

A review of WBAs by the Academy of Medical Royal Colleges in 2009 suggested widespread confusion regarding standards, methods and goals of WBAs.[30] This has prompted increased cynicism in the profession.

CONCLUSIONS

Much of the evidence that has been published already is anecdotal in nature, and there is little evidence to allow good conclusions to be drawn as to the validity and reliability of these assessments. The employment of these tools in surgical training as well as other medical disciplines emphasises the importance of ensuring that appropriate studies are carried out to confirm their reliability and validity.

Perhaps the most fundamental reason for the use of WBAs is to identify those trainees who are failing. Their misuse by both trainers and trainees can make this challenging. As a result, there remains widespread scepticism of their benefits among trainees and trainers. However, research in other specialties has demonstrated that trainees place a great deal of value on having individualised supervision and feedback. In order for these assessments to have value (rather than just exist as a tick-box exercise), the philosophy behind them needs to be embraced and they need to be used as they were originally intended: as continuous, formative assessments that help provide constructive feedback. Incorrect use of these assessments affects their reliability and therefore any validity they may have.

References

1. Phillips AW, Madhavan A. A critical evaluation of the intercollegiate surgical curriculum and comparison with its predecessor the 'Calman' curriculum. *J Surg Educ* 2013; **70:** 557–562.
2. Paice E. Is the New Deal compatible with good training? A survey of senior house officers. *Hosp Med* 1998; **59:** 72–74.
3. Scallan S. Education and the working patterns of junior doctors in the UK: a review of the literature. *Med Educ* 2003; **37:** 907–912.
4. Carr S. The Foundation Programme assessment tools: an opportunity to enhance feedback to trainees? *Postgrad Med J* 2006; **82:** 576–579.
5. Norcini J, Anderson B, Bollela V *et al*. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011; **33:** 206–214.
6. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Med Educ* 2011; **45:** 560–569.
7. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; **65:** S63–S67.
8. Postgraduate Medical Education and Training Board. *Standards For Curricula*. London: PMETB; 2005.
9. Van Der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996; **1:** 41–67.
10. Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. *J Surg Educ* 2008; **65:** 77–83.
11. Wilkinson JR, Crossley JG, Wragg A *et al*. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ* 2008; **42:** 364–373.
12. Maatsch JL, Huang R, Downing S, Barker B. *Predictive Validity of Medical Specialist Examinations. Final Report for Grant HS 02038-04, National Centre of Health Services Research*. East Lansing, MI: Michigan State University; 1983.
13. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach* 2007; **29:** 855–871.
14. Solomon DJ, Reinhart MA, Bridgham RG *et al*. An assessment of an oral examination format for evaluating clinical competence in emergency medicine. *Acad Med* 1990; **65:** S43–S44.
15. Cunnington J, Keane D. The physician review program (PREP). *Pedagogue* 1996; **7:** 1–6.
16. Southgate L, Cox J, David T *et al*. The General Medical Council's Performance Procedures: peer review of performance in the workplace. *Med Educ* 2001; **35:** 9–19.
17. Mehta F, Brown J, Shaw NJ. Do trainees value feedback in case-based discussion assessments? *Med Teach* 2013; **35:** e1166–e1172.
18. Jyothirmayi R. Case-based discussion: assessment tool or teaching aid? *Clin Oncol* 2012; **24:** 649–653.
19. Bodgener S, Tavabie A. Is there a value to case-based discussion? *Educ Prim Care* 2011; **22:** 223–228.
20. Martin JA, Regehr G, Reznick R *et al*. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; **84:** 273–278.
21. Pitts D, Rowley DI, Sher JL. Assessment of performance in orthopaedic training. *J Bone Joint Surg Br* 2005; **87:** 1,187–1,191.
22. Al Ansari A, Ali SK, Donnon T. The construct and criterion validity of the mini-CEX: a meta-analysis of the published research. *Acad Med* 2013; **88:** 413–420.
23. Brazil V, Ratcliffe L, Zhang J, Davin L. Mini-CEX as a workplace-based assessment tool for interns in an emergency department – does cost outweigh value? *Med Teach* 2012; **34:** 1,017–1,023.
24. Weller JM, Jolly B, Misur MP *et al*. Mini-clinical evaluation exercise in anaesthesia training. *Br J Anaesth* 2009; **102:** 633–641.
25. Weston PSJ, Smith CA. The use of mini-CEX in UK foundation training six years following its introduction: lessons still to be learned and the benefit of formal teaching regarding its utility. *Med Teach* 2014; **36:** 155–163.
26. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med* 2002; **77:** 900–904.
27. Pelgrim EA, Kramer AW, Mokkink HG, Van der Vleuten CP. Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study. *BMC Med Educ* 2012; **12:** 97.
28. Pereira EA, Dean BJ. British surgeons' experiences of mandatory online workplace-based assessments. *J R Soc Med* 2009; **102:** 287–293.
29. Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: are they just a tick box exercise? *Med Teach* 2011; **33:** 919–927.
30. Academy of Medical Royal Colleges. *Improving Assessment*. London: AoMRC; 2009.

This paper aimed to evaluate the use of WBAs within the surgical curriculum. As already discussed, ISCP, which is the modern incarnation of the surgical curriculum, sought to be an integrated system. This moved on from its predecessor, which was in essence just a syllabus. The new curriculum was based upon the premise of being "competency based" and in order for trainees to demonstrate their competency they were required to maintain not only a logbook of the procedures that they had performed, but also a portfolio of WBAs which would help chart their progress over their training years and serve as evidence that they had become "competent".

One of the many issues that was raised with these assessments was the subject of their reliability and validity. Validity and reliability are key components for an effective assessment tool. The evidence discussed within this paper indicated that WBAs should be used along with summative assessments so they can cover the upper two tiers of Millers competence pyramid[64] (Figure 4).

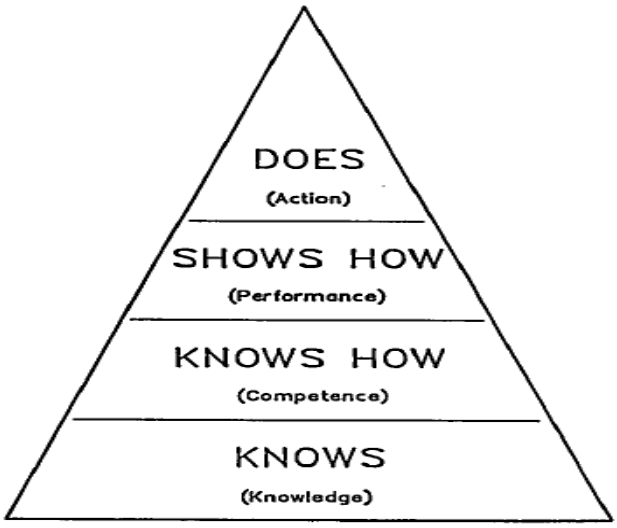

Figure 4: Miller's framework for clinical assessment[64]

Whilst this paper highlighted the anecdotal nature of much of what had been published, it also drew attention to their fundamental aim which was to help identify the struggling trainee. However, their impact is limited by engagement of both trainers and trainees.

The paucity of data available for WBAs is demonstrated by Torsney et al who sought to perform a systematic review on the same subject. They found only five observational studies that looked at the use of WBAs in a surgical setting. Further, only two of the studies

evaluated validity and reliability[65]. It was highlighted in a further letter to the editor regarding this paper that whilst many of the WBAs have face validity, they lack construct validity and are frequently misused[66]. In order for the full benefit to be obtained from these assessments "triangulation" must occur. This involves multiple assessments and multiple assessors using multiple assessment tools, in order to provide both validity and reliability.

The issue of having both validity and reliability is an important one. Whilst a tool can be deemed reliable because it will give consistent results, this in itself is not enough to validate the tool. Thus any learning tool, to be effective, must be valid i.e. measure what it is supposed to measure, and reliable- i.e. give consistent results.

3.1.5 Problems with WBAs

There are a number of potential problems with WBAs. As a relatively new tool for aiding students' education, there may still exist some doubt as to their value from both the student and educator. The majority of doctors have little or no formal instruction regarding teaching[33,34] and little information on tools available and there may be an element of distrust amongst some regarding the usefulness of WBAs.

There is also the potential for these formative assessment tools to not be used properly. Within most curricula a prescribed number of different types of assessment are usually made requisite. Whilst this helps to formalise the process, and ensure they are carried out, stating exact, or even minimum numbers of assessments may prove of little benefit. It is possible that in those jobs where a supervisor is unfamiliar or distrustful of the types of assessment employed that these formative assessments prove to be of little benefit to the student. It is more likely that a cursory approach to filling in paperwork will occur. Whilst WBAs place an onus on the trainee to ensure they occur, they also require a willing and enthusiastic trainer for benefit to be obtained. A trainee may gain more benefit from allowing their trainer to teach in a manner they are more familiar with than be forced into a contrived teaching scenario.

Within the surgical curriculum, a further problem exists in the computerised storage of assessments. While being able to keep and monitor progress on a dedicated website or programme is convenient, there is intrinsic inflexibility with only set procedures or scenarios being available, and trainers and trainees being susceptible to the vagaries of technology.

3.1.6 Surgical Trainees

Doctors are generally highly motivated individuals who have had professional development and self-directed learning indoctrinated from an early stage of training. Surgeons, are arguably an even more motivated subgroup. There must be an appropriate utilisation of training tools to meet the needs of the individuals involved. In prescribing a minimum number of WBAs for each surgical trainee to complete, a task is being set with a fixed end point. There is a danger of overlooking the educational benefit that ought to be derived from these assessments. Instead they may be regarded as part of a register that needs ticking off. Indeed there is the possible scenario that these tools are regarded as something that detracts from "proper" training and are assessments that need to be "got over with" so that one can return to focus on their job. Anecdotally, those that have just completed their yearly ARCP (Annual Review of Competency and Progression-formerly RITA (Record of in Training Assessment)), have suggested that the focus was purely on whether they had the requisite assessments signed off, rather than whether they had developed as a surgeon.

As already suggested surgical trainees are highly motivated, and almost certainly will seek out the training opportunities they think are required or have been suggested by a mentor. The need for formalising some of this is arguably not required.

3.1.7 Faculty training

The successful implementation of formative assessments requires not only the acceptance of the trainees, but also the acceptance of those teaching. Further, for maximum benefit to be gained, WBAs need to used correctly and be sensitive enough to pick up errors in student performance. Boulet et al, suggested a large variability in assessor quality when using a mini-CEX for evaluation, and this naturally translates to other WBAs[67]. The variation that is seen between assessors in WBAs is also apparent in summative assessments, explaining why viva voce exams have lost some favour. Whilst they may be discriminatory, the open ended nature of a viva exam has the potential to lead to different results for the same candidate depending on the examiner.

It has been shown that specific training in these tools leads to greater consistency between assessors[68]. The end result is that trainees are able to get feedback that is more accurate and consistent from a faculty that has been trained.

3.2 Use of Simulation in Surgical Training

3.2.1 Rationale for Simulators

The traditional method of surgical training has been along the lines of an apprenticeship, under the supervision of a surgical trainer and almost exclusively in the hospital setting. However, changes in the medical curriculum have placed new emphasis on competency based training and assessment. With doctors working patterns changing, theoretically leaving them with fewer training hours, efficient methods of developing skills are required. The exposure that each trainee gets may be quite variable and proficiency at many tasks may be difficult to attain if reliance on "what comes through the door" remains.

Thus the importance of finding a means that allows skills to be taught, learned, and developed, is vital if surgical training is to maintain a high standard with the fewer training hours that are available[69].

Surgical training is continually evolving and the use of simulators in teaching surgeons has become increasingly popular. The use of a simulation provides many benefits for the trainee, allowing task repetition and development of dexterity, without risk to patients. Further, it allows the development of well-rehearsed protocols. Simulation has long been used in the aviation industry and military to rehearse possible occurrences and make reactions second nature.

Simulation has been defined as "*A situation in which a particular set of conditions is created artificially in order to study or experience something that could occur in reality*" [70]. Thus simulation can be used to provide an environment that is safe to learn in, and can enable a tutor to provide educational support[71].

The benefit of using a simulator is that the same situation can be predictably repeated, and can be tailored to meet the educational needs of the student. The advantage of not using "real patients" also means that confounding factors such as patient stress, or embarrassment can be

eliminated, standardising the environment for students- beneficial when being used in the context of assessment.

3.2.2 Types of Simulator

There is little doubt that carrying out procedures and learning on live patients is the best method of learning. However, there are a variety of ways that "day-to- day" situations can be mimicked allowing a novice to commence learning a skill, or someone more advanced to develop their skills. Simulations have the added advantage of avoiding many ethical and safety issues that present when dealing with live patients[72]. The reproducibility of a task that can be created with these simulations means that junior surgeons can commence learning tasks and build confidence in a safe environment.

Sarker et al suggested that simulators can be broken down into several models; inorganic (subdivided into synthetic and electronic), and organic (further divided into cadaveric and animal simulators)[72].

The synthetic models can be further divided into non-life and lifelike. Non-lifelike models are simply devices that allow a particular skill to be practiced. A typical example of this is the wooden blocks used in teaching the hand tying of knots. The model itself represents no similarity to real life, however, it allows a vital skill to be taught and rehearsed. The device used on basic surgical skills courses also involves using a "shoe-lace" half black and half white. This is a useful device that allows the teacher and the student to demonstrate the knot has been tied correctly by the distinctive pattern that is produced.

Lifelike synthetic models may differ in how realistic they are. On a simple scale latex and rubber can be used to recreate surgical scenarios. Simpler models are naturally less expensive and good for teaching and learning surgical techniques whilst more expensive models can be used to recreate more complicated activities, perhaps with several steps allowing a complete procedure to be learnt. These skills can then be transferred to real life situations, with the student hopefully far nearer the top of the learning curve.

The use of electronic or virtual reality simulations is ever increasing. Virtual reality has been defined as "technology that allows people to interact efficiently with three dimensional computerised databases in real time using their natural senses and skills".[73] Simulators that are able to give a realistic feel, with tactile feedback (haptic) are particularly good for

recreating laparoscopic and endoscopic scenarios. The obvious advantage is the safe environment for learning that is recreated. Part of a procedure can be taught and practised repeatedly, or several steps of the procedure can be taught. This allows progress to be made with all the required steps and potentially a whole procedure learnt with a good degree of technical confidence before going onto operate on a live patient. The other advantages include real-time performance and reproducibility. Further the opportunity to develop a level of competence before trying these techniques on patients has ethical advantages. The ethos of see one, do one, teach one, has long been superseded and patients increasingly expect a consultant led service and thus their procedures should be carried out by the consultant. While appropriate supervision is naturally necessary until the required level of competence is reached, simulators undoubtedly will mean that levels of independence can be reached more rapidly.

Often motor skill performance is judged by time taken to perform a task and the accuracy with which the task is performed. While these two criteria are obviously important in determining whether an individual has successfully learnt how to carry out a task it fails to take into account the amount of effort required to get to the end point. Anecdotally it is often noticeable when watching different surgeons carry out the same procedure that some appear effortless whilst other can make the task appear more difficult. Whilst in some cases this may be down to experience, it could also be related to the way the individual was taught, and an element of natural ability. There is a difference between being competent at a task and effortless in the same task. In this way computerised simulators may be of use in analysing movements and effort required to accomplish a task. With the benefit of these tools individuals may not only just be able to become competent in the skill, but be able to analyse how they are completing the task and how it could be improved. Another feature would be to compare closely with the effort required by someone very accomplished in the task to identify what the fundamental differences are.

Organic models encompasses both live and cadaveric models. Within the UK use of live animal tissue to practice surgical techniques is not allowed. However, there still remain many training workshops within Europe and the US that teach surgical techniques on live animal models- for instance laparoscopic cholecystectomy on pigs. This is an expensive way of teaching and animal anatomy can vary from that found in humans. From learning a technical skill point of view, these are arguably the best after learning on real patients. Tactile feedback

should be realistic, and care in movements and how dissections are carried out are important as careless errors will be punished by bleeding or other complications.

Cadaveric models have the advantage of representing accurate anatomy and relations. Cadavers were commonplace in teaching of anatomy to undergraduate students, with regular dissection sessions. Kapadia et al found that cadaver and bench model training were equivalent and superior to text learning alone[74]. Junior doctors wishing to become a surgeon would often take jobs as demonstrators in the UK- relearning anatomy, preparing prosections and teaching the undergraduates. This provided an excellent opportunity for learning anatomy to a high level and learning careful tissue handling. However, formalin treated tissue behaves nothing like real tissue, and cannot realistically be used to learn surgical procedures. Unfortunately cadavers are being used less frequently in the undergraduate curriculum so there is less opportunity to learn dissection and gain experience in this for both undergraduate and postgraduate demonstrators.

3.2.3 Pedagogical Ideology

Teaching and learning of practical skills commonly occurs on a one-to-one basis. This is true not only in surgery, but also in anaesthetics, and those medical fields that require practical skills to be learnt such as endoscopy and angiography. Hence the traditional idea of teacher and apprentice lends itself nicely to learning to be a surgeon. Throughout their career surgeons need to continue to learn and develop. New techniques are always being pioneered and reflection upon day to day practice is essential in improving skills. This is particularly true when a procedure has "not gone to plan", or when a procedure *has* gone to plan but the desired outcome was not achieved. Part of the learning and developing process is to minimise the risk involved in carrying out a practical procedure. Simulators are an excellent way of providing a safe teaching environment which is student centred.

There are four main stages in the process to achieving competence in an area[75]. The initial stage involves the learner being *unconsciously incompetent*. At this time, the learner may have just become aware of the skill. Learners may not fully comprehend what is required to accomplish the task: however, they may feel that performing the task is within their capabilities. If an attempt is made to carry out the task the student will find that it is not as easy as they thought and they become *consciously incompetent.* This may be a shock to the student who realises that they are quite distant from being able to competently carry out the

task. As this awareness develops, formal learning occurs and experience in the procedure is gained the student may eventually attain a level of *conscious competence*. At this point the student is able to accomplish the task but has to think carefully about each step. The procedure has yet to become "natural". With more practice and experience the task may eventually become second nature, done without thinking and a level of *unconscious competence* will have been reached. At this stage the ability to carry out the skill and do something else at the same time may be possible. Arguably, at this stage the learner may be able to teach the task themselves. Although, ironically if the task has become second nature then teaching it may not be straightforward as there will be difficulty conveying how to accomplish a task that has become instinctual. Indeed, it may be that those that have reached the conscious competence stage are in an ideal position to teach. They will still have to concentrate and think about each step- usually with some clarity and can therefore pass on their understanding to students.

These steps could be related to those involved in learning to drive. One at first thinks that driving is a fairly straightforward activity; at this point one is unconsciously incompetent. As one attempts to drive an appreciation of the difficulties involved occurs and one becomes consciously incompetent. With lessons and practice the student may eventually be consciously competent, i.e. able to drive, but thought has to be placed on performing each step, depressing the clutch before changing gear and so on. Finally, all these steps will become second nature and little thought may be required in driving, the conscious competence stage. Indeed, at times confusion may occur if one stops to think about what they are doing as the process has become second nature. Simulators naturally provide an ideal opportunity for a student to move through these four phases.

Fitts and Posner suggested that motor skill acquisition involved three steps; cognition, integration and automation[76]. In the first step the student thinks about the task breaking it down to its constituent steps while attempting to understand what is technically required. This can be likened to the above stage of conscious incompetence and in the simple analogy of knot tying would be getting to grips with how to hold the suture and identifying what movements need to be made. The second step, integration, involves the student understanding what is needed to achieve the task, and being able to accomplish it to a certain degree albeit thinking about each step- for instance thinking and concentrating closely on how to move one's hands whilst tying the knot. Finally, automaticity occurs where the task is second nature and can be completed without thinking. Simulator training in laparoscopic

surgery has demonstrated that improved levels of automaticity are achieved in those at advanced stages of training and that novices are able to obtain a high level of competence. Obtaining automaticity requires substantially longer training[77].

3.2.4 Validity and Reliability of Simulation

Amongst the most important aspects of any potential simulator is its validity and reliability[78]. Validation is often thought of as incorporating verification and validation. The former being the process of ensuring whether or not the simulator operates as it was intended to. The latter involves assessing whether those conclusions attained are similar to those that would be achieved in the real-world scenario on which the simulator has been modelled. Several criterion have been made to test the validity. These include face validity, content validity, construct validity and criterion validity.

Face validity looks at how closely the simulator resembles the real-world situation and is largely an intuitive judgement. Contrastingly *content* validity relies on looking at how closely the scenario reflects what it sets out to measure i.e. the particular skill involved. *Construct validity* refers to whether the scenario correlates to a theoretical model. Outcomes should theoretically be related so that a more experienced surgeon would be expected to be do better than a novice at the skill involved. *Criterion validity* can be subdivided into *predictive validity*- how well ability in the scenario predicts future performance and *concurrent validity*- the correlation between the assessment tool and the "gold standard"[78].

Reliability is the other important measure of a simulation. This is particularly so when the simulator is also to be used as an assessment tool. Reliability refers to the degree to which the simulation yields consistent results on repeated trials. Theoretically students of the same level should perform equally well: similarly, the task should present the same level of challenge on each attempt. That is not to say the task will not appear easier as it is practiced more.

Thus in order to be a good training tool the simulator needs to be fit for purpose- actually testing the skill set that it was designed test, and also to be consistent and reproducible. Of course many simulators can be tailored so that as the student improves the task can be made more challenging.

3.3 Impact of feedback after video simulation

3.3.1 Feedback after simulation. Paper and Commentary

Nesbitt CI, Phillips AW, Searle RF, Stansby G. Randomised trial to assess the impact of supervised and unsupervised video feedback on teaching practical skills. *J Surg Educ*. 2015 Jul-Aug;72(4):697-703[4]

Randomized Trial to Assess the Effect of Supervised and Unsupervised Video Feedback on Teaching Practical Skills

Craig I. Nesbitt, MD, Alexander W. Phillips, FRCSEd (Gen Surg), FHEA,† Roger F. Searle, PhD,‡ and Gerard Stansby, FRCS**

*Northern Vascular Centre, Freeman Hospital, Newcastle upon Tyne, United Kingdom; †Northern Oesophagogastric Centre, Royal Victoria Infirmary, Newcastle upon Tyne, United Kingdom; and ‡School of Medical Sciences Education Development, Newcastle University, Newcastle upon Tyne, United Kingdom

BACKGROUND: Feedback is a vital component of the learning process; however, great variation exists in the quality, quantity, and method of delivery. Video feedback is not commonly used in the teaching of surgical skills. The aim of this trial was to evaluate the benefit of 2 types of video feedback—individualized video feedback (IVF), with the student reviewing their performance with an expert tutor, and unsupervised video-enhanced feedback (UVF), where the student reviews their own performance together with an expert teaching video—to determine if these improve performance when compared with a standard lecture feedback.

METHODS: A prospective blinded randomized control trial comparing lecture feedback with IVF and UVF was carried out. Students were scored by 2 experts directly observing the performance and 2 blinded experts using a validated pro forma. Participants were recorded on video when performing a suturing task. They then received their feedback via any of the 3 methods before being invited to repeat the task.

RESULTS: A total of 32 students were recruited between the 3 groups. There was no significant difference in suturing skill performance scores given by those directly observing the students and those blinded to the participant. There was no statistically significant difference between the 2 video feedback groups ($p = 1.000$), but there was significant improvement between standard lecture feedback and UVF ($p = 0.047$) and IVF ($p = 0.001$).

CONCLUSION: Video feedback can facilitate greater learning of clinical skills. Students can attain a similar level of surgical skills improvement with UVF as with teacher-intensive IVF. (J Surg 72:697-703. © 2015 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: feedback, clinical skills, video feedback, training, unsupervised video-feedback

COMPETENCIES: Patient Care, Medical Knowledge, Practice-Based Learning and Improvement

INTRODUCTION

Clinical feedback is defined as "specific information about the comparison between a trainee's performance and a standard, given with the intent to improve the trainee's performance."[1] The importance of feedback while learning clinical skills is well established, and it has been suggested that its absence can prevent progress.[2] However, the delivery of feedback can often be a point of criticism from students.[3] There remains much debate on the optimal method of delivering feedback. Many medical curricula have employed formative assessments to formalize and ensure that students get a regular objective appraisal.[4] There have been mixed results on the use of video feedback to improve clinical skills, with Backstein et al.[5] failing to demonstrate an improvement in orthopedic skills using video feedback. By contrast, others demonstrated a significant improvement in surgical skill acquisition following verbal video feedback.[6] A recent Best Evidence in Medical Education review of simulation commented on the importance of trainee feedback to slow learner skill decay over time.[7]

Skin suturing is a mandatory skill for all graduating doctors in the UK registering with the General Medical Council.[8] However, skin suturing is not currently taught formally within the undergraduate clinical skills curriculum at our institution, thus ensuring that participants would be novices. As skin suturing is easily recorded on video and can be assessed using established criteria, it was felt that this

Correspondence: Inquiries to Craig I. Nesbitt, MRCS, MD, Northern Vascular Centre, Freeman Hospital, Freeman Road, Newcastle upon Tyne NE7 7DN, UK; e-mail: craigiainnesbitt@gmail.com

would be the ideal clinical skill to use as the basis of this study.

Undergraduate medical students were recorded while performing a simple suturing task and then provided with a generic feedback lecture, individualized video feedback (IVF), or unsupervised video-enhanced feedback (UVF) before being asked to repeat the task. A feedback lecture is a form of generic feedback that medical students commonly receive. The 2 video feedback methods involved either students reviewing their performance with an expert providing one-to-one individual analysis of the performance (IVF) or students reviewing their own performance unsupervised but enhanced, with an expert video and a video of an expert delivering "hints and tips."

The aim of this trial was to assess the role of video-enhanced feedback (VEF), in particular to look at the potential role of UVF, in optimizing candidate performance during undergraduate medical clinical skills training.

METHODS

This prospective randomized clinical trial was carried out at Newcastle University Medical School in 2012. Ethical approval was obtained from the Newcastle University Ethics committee. Year-1 and year-2 undergraduate medical students were invited to participate in the study, and written consent was obtained from all those who volunteered. Students were assigned a unique training number (UTN) that was subsequently used for randomization and to ensure anonymity when evaluating the video performances.

The only exclusion criterion was having greater than "low novice" experience at suturing. This was defined as having performed fewer than 10 previous sutures.

Students were randomized into 3 feedback groups using a closed envelope system with students' UTNs blocked into groups of 12. This created 3 training groups of the same size. Group 1 would meet on the first day of the trial, group 2 on the second, and group 3 on the third. Students were then e-mailed confirmation of their training group date and time.

Before carrying out the suturing task, students received a short 5-minute introductory lecture explaining the task and that the procedures would be recorded and were asked to complete a questionnaire to record demographics and previous experience.

All participants were taught a basic suturing exercise using an approved Royal College of Surgeons of England technique (Intercollegiate Basic Surgical Skills). The "instrument-tied reef knot" was taught as the method for securing sutures. To ensure uniformity in teaching techniques between the 3 training groups, this teaching session was video recorded before the study, and participants watched this video on the day of the trial. The teaching video was annotated with expert demonstrations of each step and commentary explaining the technique in detail. This ensured that the teaching was entirely standardized.

After performing the task, the students received their feedback via the method they had been randomized to and were then asked to repeat the procedure (Fig. 1).

The students were scored by experts in real time and subsequently had their recordings scored by 2 further experts who were blinded to the candidate and whether the performance was prefeedback or postfeedback.

Feedback Methods

The students were randomized to any of 3 feedback methods before repeating the suturing exercise.

Group 1—Standard Lecture Feedback: 20 Minutes

The students who were randomized to the standard feedback group received a generic lecture. This feedback took place in the clinical skills laboratory and involved a 20-minute PowerPoint presentation. The presentation covered the most common errors and difficulties that had been observed in participants during a pilot experiment of suturing teaching. The lecture was delivered in a didactic fashion, and although participants were permitted to ask questions, care was taken to ensure no additional "individualized" feedback was delivered. Standard lecture feedback (SLF) was agreed to represent the closest assimilation of the feedback currently delivered by staff at Newcastle University Medical School during skills training sessions.

Group 2—UVF: 20 Minutes

The participants who were randomized to UVF were escorted to a remote private viewing room. Each participant was given a laptop computer. The computer was installed with 3 videos, which participants were instructed to watch within a 20-minute time frame. These videos included the following: a real-time unedited video of their own performance (without commentary), which lasted 7 minutes, an edited video of an expert performing the suturing exercise, with additional expert commentary, which lasted 5 minutes, and a video of an expert delivering "hints and tips," which targeted the areas that had been previously identified as causing difficulties for candidates performing this suturing exercise, which lasted 5 minutes. Additionally, 3 minutes were allowed for students to be able to stop, rewind, and replay sections of the videos as they wished.

Group 3—IVF: 20 Minutes

Participants in the IVF group watched an unedited video of their suturing performance and were given real-time, one-to-one technical skills feedback on their performance by an expert. Participants and experts were permitted to pause, rewind, and replay the video at any point and to ask technical questions. A maximum of 20 minutes was allowed for this feedback.

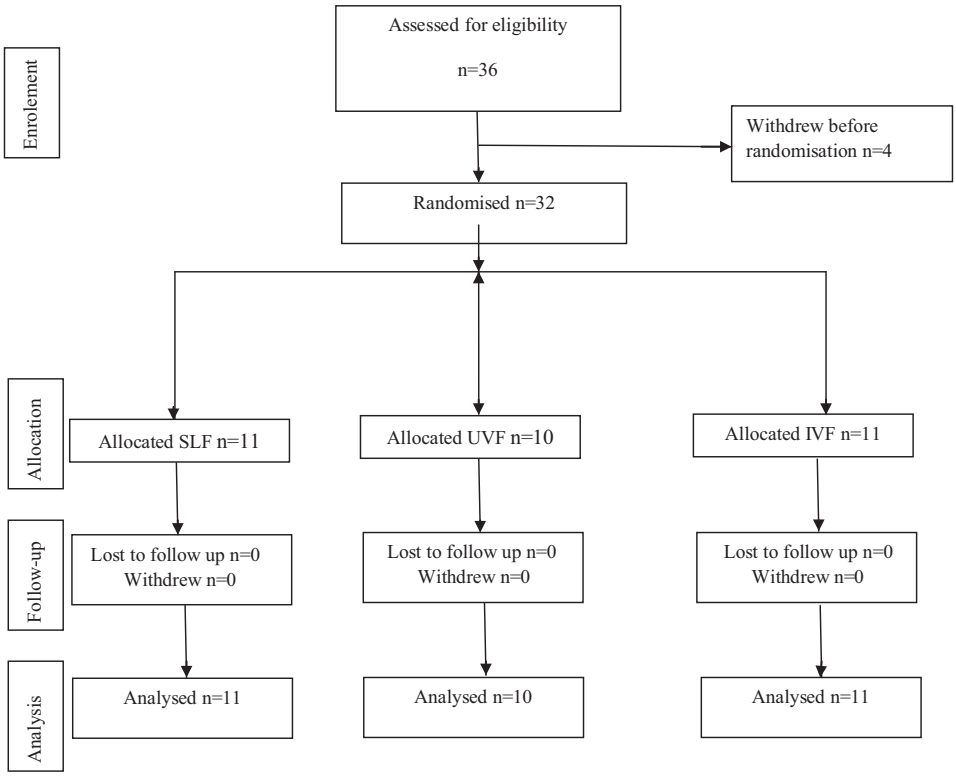

FIGURE 1. Consort diagram for the trial.

Recording of Performance

Participants' performances were recorded using a purpose-built video- and audio-enhanced training system—the Scotia Medical Observation and Training System. This involves a series of fixed-ceiling fully maneuverable cameras capable of high-definition video recording. The images are instantly stored and filed on a central computer for instant playback and assessment. The candidates were asked to suture while seated at a table. Scotia Medical Observation and Training System cameras, which had been installed at 90° to each other within each bay, were positioned so that they were focused on the synthetic skin pads. This allowed 2 views of the students' hands so that every detail was captured for subsequent analysis. Care was taken to ensure that the cameras recorded only students' hands so as to ensure the performances remained anonymous. To link candidates' performances and maintain anonymity, before each performance, a sheet identifying candidates by a UTN and highlighting if the performance was "prefeedback" or "postfeedback" was temporarily displayed into the camera.

Procedure Scoring

Students' prefeedback and postfeedback suturing performances were scored using a modified version of a previously validated clinical skills scoring tool (Objective Structured Assessment of Technical Skill).[9] This comprised both a task-specific checklist and a global rating score, providing each performance with an overall score of technical performance. Finally, examiners would indicate if they felt a participant should "pass" or "fail" that particular exercise based on their demonstrated suturing performance. Participants received a score from the experts directly observing their performance (direct observation) and one from the blinded assessors (blinded scorers).

All members of the faculty who were involved in scoring candidates on the day, met before the trial to discuss the scoring methodology. A total of 10 edited video clips of a mock candidate performing different parts of the suturing exercise, simulating varying degrees of skill, were shown to the faculty, who discussed their scoring with the intention of reducing interrater variability.

Blinded Video Analysis

All the video performances were scored by 2 experts. These experts were given complete details of the methods of the suturing experiment and viewed the training video. They were also given instructions on how to score performances using the modified Objective Structured Assessment of Technical Skill scoring tool. A practice scoring session took place during which the expert scorers discussed their scoring justification on a series of edited video clips to reduce interrater variability. The experts were completely blinded to the status of the video, whether it was prefeedback or postfeedback.

TABLE 1. Demographics

Demographic	Group 1 (SLF)	Group 2 (UVF)	Group 3 (IVF)	Significance Test of Between-Group Difference (p value)
Number	11	10	11	
Age	19.5 (17-21)	20.2 (18-24)	20.2 (19-23)	ANOVA (NS)
Sex	6 F, 5 M	5 F, 5 M	7 M, 4 F	Fisher exact test (NS)
Seniority	Year-1 students (7) Year-2 students (4)	Year-1 students (7) Year-2 students (3)	Year-1 students (6) Year-2 students (6)	Fisher exact test (NS)
Suturing experience	Low novice	Low novice	Low novice	Fisher exact test (NS)
Wear glasses	5 Yes, 6 no	5 Yes, 5 no	7 Yes, 4 no	Fisher exact test (NS)
Dominant hand	1 Left, 10 right	2 Left, 8 right	2 Left, 9 right	Fisher exact test (NS)
Play musical instrument	9 Yes, 2 no	8 Yes, 2 no	7 Yes, 4 no	Fisher exact test (NS)
Play video games regularly	3 Yes, 8 no	3 Yes, 7 no	4 Yes, 7 no	Fisher exact test (NS)
Ability to type	11 Yes, 0 no	10 Yes, 0 no	11 Yes, 0 no	Fisher exact test (NS)
Previous suturing teaching	1 Yes, 10 no	1 Yes, 9 no	2 Yes, 9 no	Fisher exact test (NS)

ANOVA, analysis of variance; NS, not significant.

Statistical Analysis

Statistical analysis was undertaken using the Statistical Package for the Social Sciences version 19 (SPSS, Chicago) and Minitab version 16. Based on a 2-tailed test, with an α level of 0.05 and power $(1 - \beta)$ of 0.8. A predicted improvement in overall procedure score by the present VEF intervention group of 30% gave a minimum of 10 subjects required in each arm. This was the same estimated percentage improvement that was used by Seymour et al.[10] The Cronbach α was used to test for the expert scorer's interrater variability. The Fisher exact test was used to compare demographic data. The Mann-Whitney *U* test was used when comparing groups' posttrial questionnaire response scores, and 1-way analysis of variance was used to compare clinical performance scores. A $p < 0.05$ was considered to be significant.

RESULTS

A total of 36 medical students indicated that they were interested in participating in the trial, with 4 students dropping out before undertaking the suturing task. The groups were demographically similar after randomization (Table 1).

Analysis of both blinded and direct observation scores showed no statistically significant difference between the 3 groups before and after feedback, indicating consistency between rating methods (Table 2).

There was a statistically significant improvement in actual score in all groups after receiving each type of feedback (Fig. 2). SLF showed an improvement from 39.3 to 50.7 ($p = 0.007$), UVF from 38.4 to 57.8 ($p = 0.003$), and IVF from 38.6 to 59.7 ($p = 0.001$).

Figure 3 demonstrates the mean improvement in overall procedure score for each of the feedback groups. This demonstrates that although all 3 groups showed improvement, there was a statistically significant improvement of both video feedback groups over SLF (group 1 vs 2 [$p = 0.047$] and group 1 vs 3 [$p = 0.001$]) but no difference in improvement between the 2 video feedback groups (group 2 vs 3 [$p = 0.595$]). The Cronbach α was 0.859, indicating good agreement between the blinded experts.

Each examiner was also asked to give the participant's performance an overall pass or fail grade. A "pass" represented a safe and satisfactory (but not necessarily perfect)

TABLE 2. Mean Values (Standard Deviation [sd]) for Direct Observation (DO) and Blinded Scorers (BS) for Prefeedback and Postfeedback Modalities

	Group 1 (SLF)		Group 2 (UVF)		Group 3 (IVF)	
	DO	BS	DO	BS	DO	BS
Task-specific checklist (TSC)						
Prefeedback	18.8 (12.11)	20.7 (12.08)	16.2 (7.63)	21.1 (10.89)	16.9 (7.75)	20.8 (8.36)
Postfeedback	29.2 (7.83)	31.2 (7.3)	33 (7.80)	35 (5.01)	35.1 (7.58)	34.9 (7.24)
Global rating score (GRS)						
Prefeedback	20.6 (6.95)	18.6 (4.08)	16.7 (4.14)	18.3 (4.54)	16.5 (4.59)	18.4 (3.7)
Postfeedback	25.6 (5.97)	22.5 (3.35)	26.2 (5.33)	23.1 (2.94)	27.2 (4.24)	24.7 (2.9)
Overall procedure score (OPS)						
Prefeedback	39.4 (39.36)	39.3 (16.09)	32.9 (11.34)	38.4 (14.77)	33.4 (11.59)	38.6 (12.44)
Postfeedback	54.7 (13.23)	50.7 (10.43)	59.1 (12.63)	57.8 (7.84)	62.6 (10.34)	59.7 (9.84)

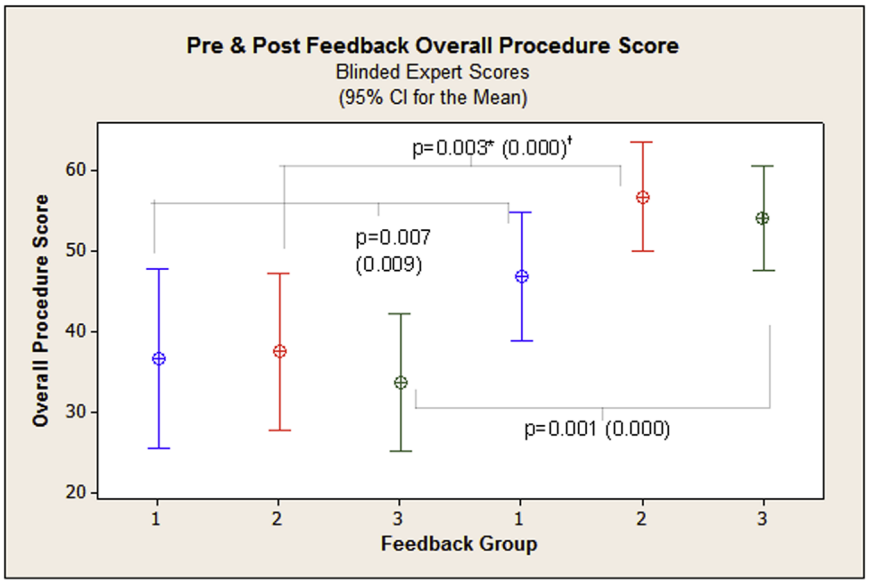

FIGURE 2. Interval plot comparing prefeedback and postfeedback OPS between SLF, UVF, and IVF. OPS, overall procedure score.

suturing performance. Performances were graded both prefeedback and postfeedback.

Table 3 summarizes the pass/fail grade for students scored by the 2 blinded expert examiners prefeedback and postfeedback. There was no statistical difference in prefeedback pass rate when comparing groups 1, 2, and 3. There was a statistically significant improvement in pass grade following both UVF ($p = 0.020$) and IVF ($p = 0.008$) but not following the feedback lecture (SLF; $p = 0.198$).

DISCUSSION

Although the results of this study demonstrate an improvement in novice suturing performance with all 3 forms of feedback, the improvement was significantly greater with both IVF and UVF. The results obtained for participants receiving IVF or UVF were similar, suggesting that the opportunity to review ones performance using video feedback plays an important role in enhancing the learning experience. The opportunity to review the task being done by an expert may also contribute to this improvement.

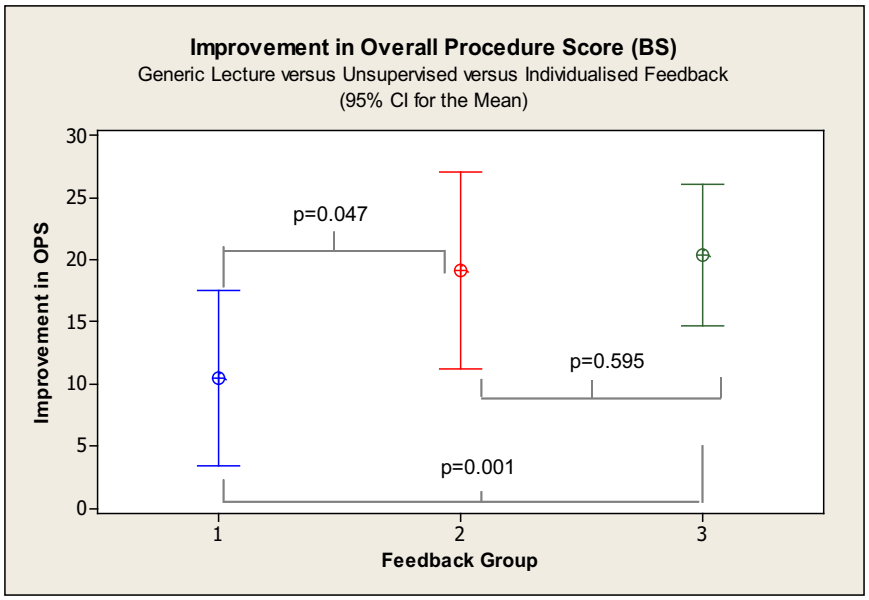

FIGURE 3. Interval plot comparing mean improvement in OPS between SLF, UVF, and IVF. OPS, overall procedure score.

TABLE 3. Pass/Fail Grade for Prefeedback and Postfeedback Performances

	Group 1 (SLF)		Group 2 (UVF)		Group 3 (IVF)	
Feedback status	Prefeedback	Postfeedback	Prefeedback	Postfeedback	Prefeedback	Postfeedback
Passes	3	7	3	9	3	10
Test of between-group difference* (p value)	0.198		0.020		0.008	

*Fisher exact test.

The use of video feedback has been established as an effective tool for improving performance both in the medical environment and in the "everyday" world. It is used extensively in both team and solo sports, where it has been shown to refine performance and permit self-critique.[11-15] In music, enhanced performance has been demonstrated in musicians who self-analyze their recorded performances, comparing these with recordings of an expert.[16] However, despite this widespread use of VEF, we are not aware of any studies to date in the music or sporting world that have compared expert one-to-one feedback with UVF, as was the strategy in this study.

Within surgical training, video analysis was initially used to identify errors during surgical procedures[17-19] and subsequently to enhance resident training in plastic surgery, thus allowing self-scrutiny.[20]

Feedback plays an important role in mastering surgical technical skills. During the *associative* phase of technical skills acquisition in surgery, the learner is practising and comparing their performance with that of an expert. It is during this phase that feedback is crucial.[21] However, despite this level of understanding, the exact role of feedback during technical skills training in medicine and surgery remains an area of debate. Rogers et al. demonstrated superior performance in surgical novices' performing a simple knot-tying exercise following a traditional lecture-and-expert-feedback technique compared with a novel computer-assisted training package. They concluded that expert feedback is essential to maximize candidate improvement.[22] However, others compared no feedback, video-assisted expert feedback, and video with self-review (which included the trainee watching their own performance with no feedback at all) in orthopedic trainees and demonstrated no improvement in performance in either feedback group when compared with the no feedback control.[23]

Feedback in any training domain is said to be either internal or external.[24] Internal feedback is generated by the learners as they compare their performance with that of an expert. On the contrary, external feedback is directed by the trainer who critiques the trainee, pointing out errors and strategies to improve their performance. Rogers et al.[24] concluded that the lack of external feedback when candidates used a computer-assisted learning package caused the inferior improvement seen in their group than with expert feedback. The results of the present trial suggest that a form of external feedback can be assimilated through the use of "expert" video recordings. This could explain the lack of significant difference in clinical performance scores when comparing the UVF and IVF groups.

Although the current study has a relatively low sample size, a significant difference in performance between VEF and the generic feedback lecture was established. It is acknowledged that the study design omitted control groups that received either no feedback (to define the baseline of how much learning occurs by repetition) or which received a video of their own performance with no type of accompanying teaching, as these were considered to be unfavorable on educational grounds. The study was not powered to look at whether a statistical difference existed between the 2 video groups. However, if a 30% difference, which was determined as "clinically significant" had existed, it should have been observed. More research looking at whether there is any difference in "real terms" between the 2 video feedback groups would be valuable.

Further investigation needs to be carried out to determine if these findings regarding the value of UVF demonstrate consolidated improvement of performance following a time delay. This was beyond the practical capabilities of the current study. It is also necessary to determine if UVF can be extrapolated to "nontechnical skills," such as clinical examination procedures or communication skills. Establishing that UVF can be as effective as individual feedback has extensive ramifications in costs and resources. The implications of UVF for improving the students' clinical skills learning experience are potentially immense and warrant further investigation.

REFERENCES

1. Van de Ridder JMM, Stokking KM, McGaghie WC, ten Cate OTJ. What is feedback in clinical education? *Med Educ*. 2008;42(2):189-197. http://dx.doi.org/10.1111/j.1365-2923.2007.02973.x.
2. Mahmood T, Darzi A. The learning curve for a colonoscopy simulator in the absence of any feedback: no feedback, no learning. *Surg Endosc*. 2004;18(8):1224-1230. http://dx.doi.org/10.1007/s00464-003-9143-4.
3. National Student Survey. *Findings and trends 2006-2010*. Available at: ⟨www.hefce.ac.uk/media/hefce/content/pubs/2011/201111/11_11.pdf⟩; Accessed 24.05.12.

4. Phillips AW, Madhavan A. A critical evaluation of the intercollegiate surgical curriculum and comparison with its predecessor the "Calman" curriculum. *J Surg Educ*. 2013;70(5):557-562. http://dx.doi.org/10.1016/j.jsurg.2013.03.003.

5. Backstein D, Agnidis Z, Sadhu R, MacRae H. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Can J Surg*. 2005;48(3):195-200.

6. Farquharson AL, Cresswell AC, Beard JD, Chan P. Randomized trial of the effect of video feedback on the acquisition of surgical skills. *Br J Surg*. 2013;100(11):1448-1453. http://dx.doi.org/10.1002/bjs.9237.

7. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 2005;27(1):10-28. http://dx.doi.org/10.1080/01421590500046924.

8. General Medical Council. *Tomorrow's Doctors: Duties of a Doctor Registered With the General Medical Council*. 2009:77-81.

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

10. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. 2002;236(4):458-463. http://dx.doi.org/10.1097/01.SLA.0000028969.51489.B4 [discussion 463-464].

11. Lee A, Keh N, Magill R. Instructional effects of teacher feedback in physical education. *J Teach Phys Educ*. 1993;12:228-243.

12. Christina R, Barresi J, Shaffner P. The development of response selection accuracy in a football linebacker using video training. *Sport Psychol*. 1990;4:11-17.

13. Winfrey M, Weeks D. Effects of self-modelling on self efficacy and balance beam performance. *Percept Mot Skills*. 1993;77:907-913.

14. Caliendo E, Kopacz R. Improving Student Motivation and Performance in Music Programs; 1999.

15. Winter H, Rees J. Gentamicin-containing collagen implants for prophylaxis and treatment of surgical site infection: an efficacy review. *Int J Surg*. 2012;10(suppl 1):S1-S28.

16. Caliendo E, Kopacz R. Improving student motivation and performance in music programs. Masters action research project. Saint Xavier University and IRI/Skylight; 1999.

17. Goldman LI, Maier WP, Rosemond GP, Saltzman SW, Cramer LM. Teaching surgical technique by the critical review of videotaped performance—the surgical instant replay. *Surgery*. 1969;66(1):237-241.

18. Goldman LI, Maier WP, Saltzman SW, Rosemond GP. Television equipment and its application in the learning of surgical skills. *Curr Top Surg Res*. 1970;2:545.

19. Goldman LI, Saltzman SW, Rosemond GP. Television equipment and its application in the learning of surgical skills. *J Med Educ*. 1972;47(10):786-788.

20. Stranc MF, McDiarmid JG, Stranc LC. Video assessment of surgical technique. *Br J Plast Surg*. 1991;44(1):65-68.

21. Kopta JA. An approach to the evaluation of operative skills. *Surgery*. 1971;70(2):297-303.

22. Rogers DA, Regehr G, Yeh KA, Howdieshell TR. Computer-assisted learning versus a lecture and feedback seminar for teaching a basic surgical technical skill. *Am J Surg*. 1998;175(6):508-510.

23. Backstein D, Agnidis Z, Regehr G, Reznick R. The effectiveness of video feedback in the acquisition of orthopedic technical skills. *Am J Surg*. 2004;187(3):427-432. http://dx.doi.org/10.1016/j.amjsurg.2003.12.011.

24. Rogers DA, Regehr G, Howdieshell TR, Yeh KA, Palm E. The impact of external feedback on computer-assisted learning for surgical technical skill training. *Am J Surg*. 2000;179(4):341-343.

Simulators are likely to become increasingly important in the future. This is due to the changes that have already been discussed, and the additional pressures on resources and time for training. Whilst some simulators have the capacity for analysing what you have done and may look at components such as efficiency of movement or speed of performing a task, they will not necessarily be able to critically evaluate what is being done by an individual and how this can be improved.

This pilot study sought to determine the impact of three types of feedback on learning a simple suturing skill. Participation was purely voluntary and novices (medical students) were given a generic lecture on how to perform a simple suturing task. They then were recorded performing the skill and were randomly divided into three groups receiving three types of feedback; 1) the first cohort received a generic lecture, 2) the second cohort were provided with footage of their performance with an accompanying expert video, and hints and tips– so candidates could carry out self-criticism and hopefully improve, 3) the final cohort sat with an expert and went through the initial performance. Both the latter groups did significantly better after teaching than the group receiving just a lecture as feedback. However, there was no difference between the latter two cohorts.

This paper raises a number of interesting questions. First, will these findings be applicable to some tasks and not others? Second, will there be any difference in ability to self-criticise according to the experience of the trainee? Third, will the findings be applicable to non-technical skills? Fourth, will there be a difference in skill attrition over time depending on feedback type received?

There are a number of shortcomings within this study- first the small sample size may introduce a degree of bias, and the task performed is very basic. But, this is a task that undergraduates (and even postgraduates) may have to master and providing an easy low resource mechanism for feedback could expedite learning. This is an introductory study into this type of learning and may allow basic tasks to be mastered without requiring much feedback from tutors. This study also did not allow for specific practice: however, this was done deliberately. Whilst practice may make a person more proficient it is likely to ingrain technical errors. The aim was to see if these technical errors could be diminished with the types of feedback provided.

In this study, the lecture feedback group acted as a control group, as it was felt that it was not educationally ethical to have a group that received no feedback. It would have been

interesting to see what impact simple repetition of the task would have had, but reasonable to expect that it would be no more efficacious (or detrimental) than those repeating the task after a standardised lecture.

This study served to demonstrate that novice medical students were able to self-critique in a simple technical skill and significantly improve their performance- and importantly that having an expert tutor giving individual feedback for such a task conferred no significant difference in improvement.

3.3.2 Expert feedback in simulation training, versus self-directed learning. Paper and Commentary

Phillips AW, Bookless LR, Matthan J, Whitehead I, Madhavan A, Nesbitt CI, Stansby G. Individualised expert feedback is not essential for improving clinical skills performance in novice learners: A Randomised Trial. *J Surg Ed* 2017; 74:612-620[5]

Individualised Expert Feedback is Not Essential for Improving Basic Clinical Skills Performance in Novice Learners: A Randomized Trial ☆

Alexander W. Phillips, MA, FRCSEd, FFSTEd, Joanna Matthan, MA, MBBS,† Lucy R. Bookless, MBChB, MRCSEd,‡ Ian J. Whitehead, MBChB, MRCS,§ Anantha Madhavan, MBChB, MRCSEd,* Paul Rodham, MRes, MBBS,‖ Anna L.R. Porter, MRes, MBBS,‖ Craig I. Nesbitt, MD, MRCS,¶ and Gerard Stansby, MA, FRCS*

*Northern Oesophagogastric Cancer Unit, Royal Victoria Infirmary, Newcastle upon Tyne, United Kingdom; †Anatomy and Clinical Skills Department, School of Medical Education, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; ‡Department of General Surgery, James Cook University Hospital, Middlesbrough, United Kingdom; §Department of General Surgery, St Helen's and Knowsley Hospitals NHS Trust, St Helens, United Kingdom; ‖Royal Victoria Infirmary, The Newcastle upon Tyne NHS Foundation Trust, Newcastle upon Tyne, United Kingdom; ¶Department of Vascular Surgery, James Cook University Hospital, Middlesbrough, United Kingdom; and Department of Vascular Surgery, Freeman Hospital, Newcastle upon Tyne, United Kingdom

OBJECTIVE: To determine whether unsupervised video feedback (UVF) is as effective as direct expert feedback (DEF) in improving clinical skills performance for medical students learning basic surgical skills—intravenous cannulation, catheterization, and suturing.

BACKGROUND: Feedback is a vital component of the learning process, yet great variation persists in its quality, quantity, and methods of delivery. The use of video technology to assist in the provision of feedback has been adopted increasingly.

METHODS: A prospective, blinded randomized trial comparing DEF, an expert reviewing students' performances with subsequent improvement suggestions, and UVF, students reviewing their own performance with an expert teaching video, was carried out. Medical students received an initial teaching lecture on intravenous cannulation, catheterization, and suturing and were then recorded performing the task. They subsequently received either DEF or UVF before reperforming the task. Students' recordings were additionally scored by 2 blinded experts using a validated proforma.

RESULTS: A total of 71 medical students were recruited. Cannulation scores improved 4.3% with DEF and 9.5% with UVF ($p = 0.044$), catheterization scores improved 8.7% with DEF and 8.9% with UVF ($p = 0.96$), and suturing improved 15.6% with DEF and 13.2% with UVF ($p = 0.54$). Improvement from baseline scores was significant in all cases ($p < 0.05$).

CONCLUSION: Video-assisted feedback allows a significant improvement in clinical skills for novices. No significant additional benefit was demonstrated from DEF, and a similar improvement can be obtained using a generic expert video and allowing students to review their own performance. This could have significant implications for the design and delivery of such training. (J Surg Ed 74:612-620. © 2017 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: simulation, clinical skills, feedback, video technology, practical clinical skills, technology-enhanced learning

COMPETENCIES: Practice-Based Learning and Improvement

INTRODUCTION

The importance of feedback while learning clinical skills is well established.[1] Both undergraduate and postgraduate students

☆The authors would like to express their gratitude to Vygon who provided the cannulas for this study and Bard who provided the catheters.

Correspondence: Inquiries to Alexander W. Phillips, MA, FRCSEd, FFSTEd, Northern Oesophagogastric Cancer Unit, Royal Victoria Infirmary, Victoria Road, Newcastle upon Tyne NE1 4LP, UK; e-mail: awphillips@doctors.net.uk

  1931-7204/$30.00 http://dx.doi.org/10.1016/j.jsurg.2016.12.003

frequently complain that a major fault in their training lies with the quantity and quality of feedback provided.[2-4]

The use of simulation for teaching clinical skills has gained increasing popularity over the past few years. The reasons for this are multifaceted. Training is delivered over a shorter period of time and in fewer hours than ever before,[5] resulting in the need for more efficient training.[6] The ability of the students to learn and practice skills in a safe and reproducible environment before having to perform the skill in the real world has been shown to be invaluable economically, medicolegally, and ethically, and improves patient comfort and safety.[7-9] These simulated environments also provide an excellent opportunity for students to have their performances recorded so that they can review, reflect, and learn from them. Previous studies have demonstrated that this review and reflection can lead to a significant improvement on the subsequent ability to perform the task involved.[10] It has also been previously suggested that students best acquire psychomotor skills when they are self-guided in their approach to feedback and when they set independent goals.[11] Recent studies have indicated that giving students the opportunity to "self-regulate" their learning, defined as being able to generate, plan, and adapt actions to attain goals,[12] improves learning in a simulated environment.[13-15]

A recent study at Newcastle University showed that video-assisted feedback contributed to a significant improvement in medical students' ability to perform a basic suturing task.[16] The ability to review one's recorded performance, with or without an expert, led to a significant improvement over simply repeating the skill after having had generic feedback.

Intravenous cannulation, urethral catheterization, and simple skin suturing are all key clinical skills that medical students are expected to be competent at performing by the end of their medical training.[17] However, anecdotally, it can be difficult for students to get feedback on their performance of these tasks to determine whether they are making adequate progress. These 3 skills are all easily recorded and high-quality simulators are available for students to practice the skills on in the safety of a simulated environment before they need to perform these tasks on patients.

Determining whether there is a difference in impact between these 2 feedback types could have a significant effect on the ability to deliver a wide range of clinical skills simultaneously to a larger number of students, while enabling the students to fulfill their desire for autonomy within their learning and give them the skills to continue life-long learning.[18] The aim of this study was to evaluate whether unsupervised video feedback (UVF) has a comparable effect to direct expert feedback (DEF) with respect to improvement in clinical performance for 3 key basic tasks—intravenous cannulation, urethral catheterization, and simple skin suturing.

METHODS

This was a prospective randomized trial carried out at Newcastle University in 2015. All medical students were invited to participate, and written consent was obtained. Full ethical approval was obtained from the Newcastle University Ethics Committee.

Students were given an introductory generic prerecorded teaching lecture before performing each skill and then undertook the skill with their attempt recorded. They were subsequently given feedback according to the group they had been randomized into, DEF or UVF. Students performed all 3 skills (intravenous cannulation, catheterization, and suturing) and were randomized, in no particular order, to receive DEF with 2 skills and UVF with 1, or UVF with 2 skills and DEF with 1.

Technology and Feedback Mechanisms

Students performed 2 attempts at each skill, 1 immediately after receiving a generic lecture, with narration on best practice, on how to perform that skill, and a second attempt after receiving feedback. The initial lecture for each skill was prerecorded and presented as a video in a learning space in order to standardize initial teaching.

Videos of each student's performance were made using the Scotia Medical Observation and Training System (SMOTS), a purpose-built video- and audio-enhanced recording system, installed for training purposes in a variety of clinical and nonclinical environments worldwide. A series of fixed-ceiling, fully maneuverable cameras capable of high-definition video recording enable recordings to be stored and filed on a central computer for instant video playback and assessment. After each student's performance, videos were uploaded to a computer for them to review.

Students randomized to receive DEF had 20 minutes to review their unedited performance accompanied by an expert (a postgraduate clinician, and signed off as competent at each skill). Participants were permitted to pause, rewind, and replay the performance with the expert giving tips on improvement or pointing out where errors were made. Experts had met before the trial and reviewed sample videos in order to standardize what was agreed as the best technique and to help ensure the same level of feedback was provided by each trainer.

Students randomized to receive UVF also had 20 minutes to review their unedited performance. In addition, they had access to an expert video with commentary that they could watch once they had reviewed their own performance. They were able to rewind and replay both of these videos over the time period allocated for feedback.

Clinical Skills Performed

Three basic clinical skills, regarded as essential,[19] were selected for this trial. Students were taught according to the agreed University or Royal College of Surgeons guidelines.

(1) *Venous cannulation*: Students were taught and then asked to perform intravenous cannulation on an

artificial teaching model. There was particular focus on use of the aseptic technique following the accepted guidelines currently employed by Newcastle Medical School.

(2) *Male Catheterization*: Students were taught and then asked to insert a urinary catheter in an artificial teaching model. There was particular focus on the use of the aseptic technique.

(3) *Simple skin suturing*: Students were taught to perform an "instrument tied reef knot" as described on the Royal College of Surgeons Basic Surgical Skills course. This was performed on an artificial skin model using a 2/0 polyfilament suture.

Exclusion Criteria

As the focus of this trial was to establish the effect of a particular type of feedback on the subsequent performance of a practical clinical skill, it was necessary to exclude (1) nonmedical students (e.g., biomedical sciences students) as well as (2) any student who, at the outset of the trial, already possessed greater than novice experience at any of the skills taught and assessed. A novice was defined as someone with less than 10 previous attempts at the skill.

Randomization

Students were allocated a unique identification number to allow anonymity and then block-randomized to ensure that 2 approximately equal groups were obtained. Videos were edited by an independent person who removed, as far as possible, components that would allow assessors to identify participants (Fig. 1).

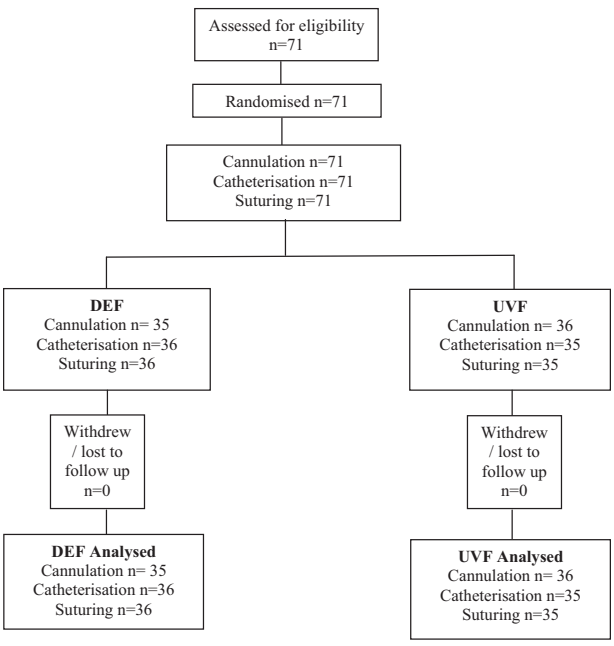

FIGURE 1. CONSORT diagram for study.

Assessment

Videos were anonymized and marked by 2 independent assessors. Three assessors were used in total, and all the 3 were higher surgical trainees (postgraduate year 5 and above). Performances were scored using a task-specific checklist and a global ratings score of *pass*, *borderline pass,* or *fail*. Assessors initially scored 10 "mock" videos together with discussion to establish consistency and reduce interrater variability. Assessors were completely blinded to the status of the video, to whether it was a prefeedback or postfeedback performance and to what type of feedback had been used. Checklist construct validity was established by a pretrial pilot with blinded assessors marking undergraduate and postgraduate performances that determined improved scores in more experienced candidates.

Statistical Analysis

Statistical analysis was carried out using the Statistical Package for the Social Sciences version 20 (SPSS, Chicago). Checklist scores were analyzed using the Shapiro-Wilkes test to confirm that data were normally distributed. Intraclass correlation coefficient (ICC) was used to test for the expert scorer's interrater variability. The Fisher exact and chi-squared test was used to compare demographic data. Student's *t*-test was used to compare improvement between DEF and UVF for each skill and Mann-Whitney *U* test for global scores. A $p < 0.05$ was considered to be significant.

Power Calculation

A power calculation was carried out based on a previous pilot study that evaluated both these feedback techniques with suturing performance.[16] The aim of this study was to compare improvement between the DEF group and the UVF group sensitive to a difference of 10% between cohorts. Based on an α level of 0.05 and power $(1-\beta)$ of 0.80, a minimum of 23 subjects in each cohort was required. A total of 71 students were recruited to allow for possible participant loss.

RESULTS

A total of 71 medical students were recruited to the trial. All completed the 3 clinical tasks twice, receiving one of the designated types of feedback between attempts. All candidates received each form of feedback on at least 1 occasion. Candidates' videos were scored by 2 blinded assessors and mean scores were used for analysis.

There was good correlation of checklist scores (*pass*, *borderline pass*, and *fail*) with the global scores (Figs. 2-4).

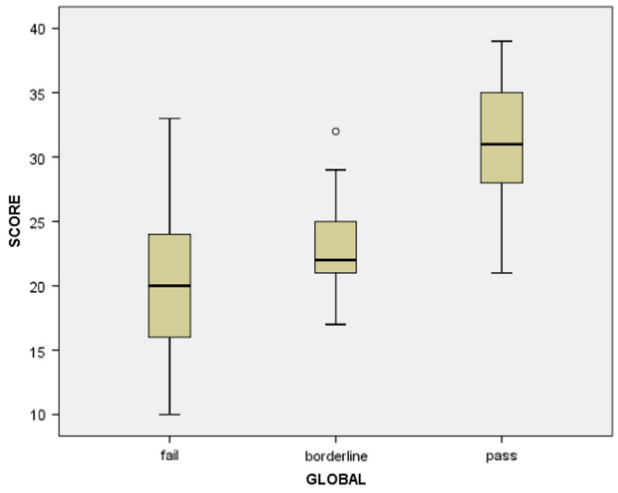

FIGURE 2. Box and whisker plot of Global scores in relation to checklist score for cannulation.

There was a significant difference in checklist scores when grouped by global category for all 3 skills ($p < 0.001$ in all cases).

Intravenous Cannulation

Checklist Scores

Students were randomized in a manner where 35 students received DEF and 36 received UVF. Baseline demographics were similar between groups, as was experience in performing the task (Table 1). Candidates were scored out of 40. There was a statistically significant improvement in scores with both DEF and UVF ($p < 0.05$). Students receiving DEF improved by a mean of 1.7 (standard deviation [SD] = 4.5) points (4.25%), and those receiving UVF improved by 3.8 (SD = 3.67) points (9.5%). There was a statistical difference in the level of improvement between DEF and UVF ($p = 0.044$). Assessor ICC was 0.94.

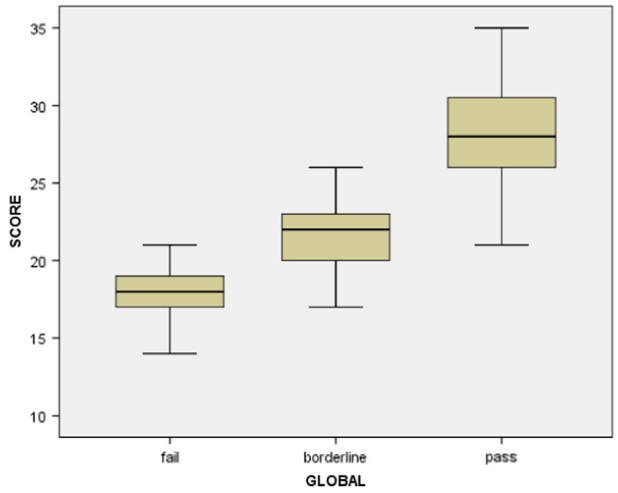

FIGURE 3. Box and whisker plot of Global scores in relation to checklist score for catheterization.

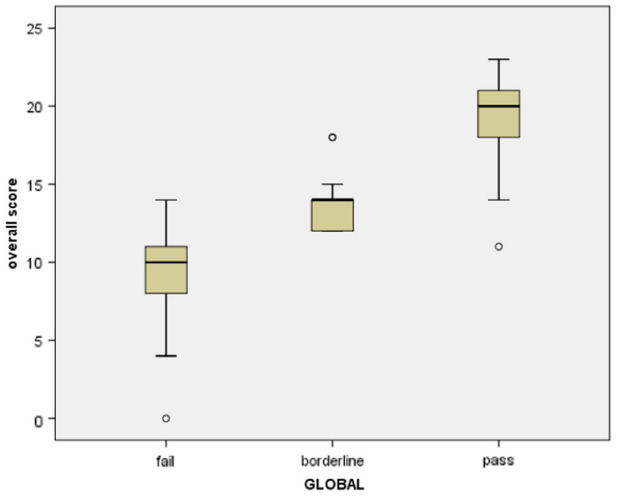

FIGURE 4. Box and whisker plot of Global scores in relation to checklist score for suturing.

Global Scores

The 2 assessors were asked to give an overall global impression of each video marked. This was to be graded as *pass*, *borderline pass,* or *fail*. Of the 142 videos marked, 125 (88%) were marked in agreement with regard to *pass*, *borderline pass,* and *fail*. Of the 17 videos where a discrepancy existed, only 5 (3.5%) had 1 marker scoring a clear *pass* while the other a clear *fail*. In the remaining 12 videos, 1 assessor scored a *fail* while the other scored a *borderline pass*.

In the DEF group, the *pass* rate increased from 54% to 83 % (assessor 1) and 71% to 94% (assessor 2). In the UVF group, the *pass* rate increased from 53% to 81% (assessor 1) and 58% to 89% (assessor 2). There was no significant difference in improvement between groups ($p = 0.54$) (Table 2).

Catheterization

Checklist Scores

Students were randomized in a manner where 36 students received DEF and 35 received UVF. Baseline demographics were similar between groups, as was the experience in performing the task (Table 1). Candidates were scored out of 27. Students receiving DEF improved by a mean of 2.36 (SD = 3.3) points (8.7%), and those receiving UVF by 2.4 points (SD = 3.0) (8.9%). There was no significant difference in level of improvement between the cohorts ($p = 0.96$). Assessor ICC was 0.91.

Global Scores

Both assessors were asked to provide an overall global impression of each video marked as *pass*, *borderline pass,* or *fail*. Of the 142 videos marked, 108 (76%) were marked

TABLE 1. Baseline Demographics for Each Group

	Cannulation			Catheterization			Suturing		
	DEF	UVF	p	DEF	UVF	p	DEF	UVF	p
n	35	36		36	35		36	35	
Age	21 (18-28)	22 (18-37)	0.789	22 (18-33)	22 (18-37)	0.181	22 (18-37)	22 (18-33)	0.334
Male:Female	17:18	18:18	0.92	17:19	18:17	0.92	19:17	16:19	0.72
Medical school year									
1	8	9		9	8		8	9	
2	5	5		5	2		6	4	
3	12	5	0.26	8	9	0.76	6	11	0.53
4	5	10		7	8		9	6	
5	2	4		4	2		4	2	

in agreement with regard to *pass*, *borderline pass,* and *fail*. Of the 34 videos where a discrepancy existed, only 10 (7%) had 1 marker scoring a clear *pass* while the other a clear *fail*. In the remaining 24 videos, 1 assessor scored a *fail* while the other scored a *borderline pass*.

In the DEF group, the *pass* rate increased from 72% to 83 % (assessor 1) and 64% to 81% (assessor 2). In the UVF group, the *pass* rate increased from 57% to 83% (assessor 1) and 54% to 63% (assessor 2). There was no significant difference between DEF and UVF (p = 0.78) (Table 2).

Suturing

Checklist Scores

Students were randomized in a manner where 36 students received DEF and 35 received UVF. Baseline demographics were similar between groups, as was the experience in performing the task (Table 1). Candidates were scored out of 24. Students receiving DEF improved by a mean of 3.75 points (15.6%) (SD = 4.5), and those receiving UVF improved by 3.17 points (SD = 3.4) (13.2%), (p = 0.54). Assessor ICC was 0.856.

Global Scores

The global scores from assessors produced an agreement with regard to *pass*, *borderline pass,* or *fail* in 128 (90%) videos. Of the 14 videos where a discrepancy existed, there were 9 episodes (6.3%) where 1 marker indicated a clear *pass* and the other a clear *fail*, and on 5 occasions (3.5%) 1 marker indicated a *fail* and the other a *borderline pass*.

In the DEF group, the *pass* rate increased from 89% to 100% (assessor 1) and 75% to 100% (assessor 2) compared to the UVF group where the global *pass* rate improved from 77% to 94% (assessor 1) and from 80% to 100 % (assessor 2) (p = 0.33) (Table 2).

DISCUSSION

This study showed comparable improvement in performances between DEF and UVF. Indeed, for 2 of the 3 skills, UVF led to a greater improvement in scores than DEF, which was statistically significant using the checklist score for the intravenous cannulation skill. Global outcomes improved with both types of feedback in each clinical skill, but there was, again, no significant difference in improvement between feedback mechanisms.

The use of global ratings has been increasingly advocated as better able to discriminate between performances.[20-23] In this study, there was good agreement between assessors in checklist and global scores, and also concordance between the 2 marking mechanisms. Internal consistency was established by high ICC scores with the checklist, suggesting that each candidate was scored similarly between assessors, which add to the validity of the score systems.[24] Although global scores may be beneficial in summative assessments, a checklist score has the benefit of helping identify particular items in a skill that are poorly done either by individuals or the group. Both scoring systems were employed to ensure more robust results.

TABLE 2. Global Scores From Each Assessor

	DEF				UVF				
Skill Tested	Mean Checklist Improvement (CI)	Global Change			Mean Checklist Improvement (CI)	Global Change			p
		Pass	Borderline	Fail		Pass	Borderline	Fail	
Cannulation	1.77 (0.22-3.33)	12	7	−19	3.78 (2.54-5.02)	15	6	−21	0.044
Catheterization	2.35 (1.24-3.48)	16	−6	−10	2.40 (1.36-3.44)	10	3	−13	0.96
Suturing	3.75 (2.24-5.26)	15	−2	−13	3.17 (2.00-4.34)	18	−5	−13	0.54

The purpose of this study was to evaluate these 2 feedback mechanisms on basic clinical skill performance. The use of simulation has been increasingly integrated into both undergraduate and postgraduate curricula.[25-27] Simulation allows individuals to learn and practice a skill in a safe manner without the concerns of patient discomfort.[12] However, irrespective of whether skills are learnt on "real" patients or in a simulated environment, feedback—and how it is delivered—remains a vital component of the learning pathway. Provision of effective feedback places heavy demand on faculty time and designs that can minimize this need to be sought.[12] UVF reduces the need for faculty review and could change the way basic clinical skills are taught, providing sufficiently high-quality teaching videos that are created to facilitate self-direction. It is difficult to determine how far this learning mechanism can be employed when learning clinical skills and surgical skills. A number of factors may influence the potential effect. These include the quality of the videos, the skill being learned, and the experience of the student and the initial teaching experience of the student. High-quality videos that clearly demonstrate the steps, possibly with appropriate narration, need to be created as a baseline resource for this form of teaching to be effective. Similarly, not all skills may lend themselves to being "taught" using video feedback. However, it would appear that, for basic clinical skills, there is some positive effect. More advanced clinical and surgical skills may potentially need to be broken down into simpler steps, which would then allow UVF to play a more meaningful role. Finally, the experience of the learner, not just at the skill they are trying to learn, but with respect to stage of learning, whether they have been taught and developed the skill of appropriate self-criticism as well as their own inherent temperament, may influence the effectiveness of UVF.

Video feedback is used through a number of disciplines from sport to music, as well as in learning practical skills, and has been shown to help those learning a skill improve technique and, ultimately, performance.[28–30] Within these areas, previous studies have usually sought to utilize an expert to review the performance of the student using video technology.

Receiving feedback from an expert could arguably be regarded as the gold standard of feedback provision, as it allows candidates to have individualized instruction as to where errors have been made and how a skill can be improved. However, the ability to critically appraise one's own performance is an equally important skill that should be honed at an early stage and, when done correctly, can have a significant effect on performance. Immediate feedback has been advocated by some as necessary for the mastery of procedures.[31] However, other studies have suggested that although immediate feedback can affect positively at the time of learning, it has no benefit on long-term skill retention and could even be detrimental to actual skill *learning*.[32] This performance-learning paradox, whereby immediate performance improvement is not reproduced in future skills retention,[33] needs further investigation with regard to the effect of UVF. There is good reason to be optimistic that this will not be the case. The detrimental effect of immediate feedback on long-term learning has been attributed to students relying on the tutor for feedback on improving their immediate performance and, consequently, not engaging their own discriminating capabilities, so in the future learning situations, without the support of a trainer, their performance falls.[32,34] UVF encourages reflection and self-directed learning when it is combined with provision of excellent quality teaching material. Further, its effect is easily reproducible, as videos can be made available and, indeed, there is the potential with technology for students to develop a portfolio of clinical skill attempts. The ability to record and access a performance means that students are able to practice a skill and then evaluate how they have done. Although the present study confirms that both types of video feedback result in improved performance, further evidence is required to determine if actual *learning* is improved. Learning implies that the skill can be performed at a later date to the same level. Access to one's performance may enhance practice and, consequently, skills learning.

The pilot study for this trial revealed that both these video feedback techniques were superior to a standardised lecture and repetition of suturing tasks.[16] The skills performed in this study are straightforward, basic tasks that medical students would be expected to have mastered on completion of their training. They provide the building blocks for performing more advanced procedures and, thus, finding effective ways of improving how they can be taught and learnt is of importance. Further, it may be possible that the effect of UVF on performance can be translated to more advanced clinical and surgical skills. Video-based coaching is increasingly used in the learning and teaching of clinical skills and, within surgery, is particularly advocated in minimally invasive techniques.[35-37] Strandbygaard et al evaluated the effect of instructor feedback in a simulated laparoscopic salpingectomy. Their findings demonstrated instructor feedback–expedited achievement of proficiency; however, interestingly, most of their control group, which was allowed performance self-review, attained proficiency and eventually achieved higher performance levels.[38] A systematic review of simulated laparoscopic training showed that the presence of an instructor is not necessarily associated with any advantage over self-review with regard to time taken to perform a skill or scored outcomes on performances.[39]

There are a number of limitations with the present study. First, the skills performed were relatively basic clinical skills, and it would be of interest to try and determine the effect of video feedback on more technically demanding procedures. It may be that there is a threshold that exists for self-learning

before an expert's objective feedback becomes invaluable. However, it may also be true that, as one becomes more experienced with clinical skills, one is better able to effectively self-evaluate performances.

Second, there was no longitudinal evaluation of skill retention. Further study into whether one type of feedback leads to better skills retention is required. However, it should be noted that it is straightforward for students to keep a recording of their performance and review this themselves at regular intervals as a method of reminding themselves of how to perform a skill, as well as a means of reviewing the progress they are making. Therefore, UVF may lead to performances being sustained at the same level. This is an important area for future study as it relates to the actual task of *learning*, as opposed to simple short-term performance improvement.

Thirdly, there was no control group evaluating the effect of simple task repetition. A previous study demonstrated that task repetition after generic feedback did lead to performance improvement in suturing but was significantly worse than the 2 types of feedback evaluated in this study.[16] It was felt that including a third group that received no feedback and simply performed the task a second time was not educationally "ethical" and unacceptable to the student body and, therefore, not included in the present study format.

Finally, it is difficult to ensure that the quality of feedback provided in the DEF group was consistent across trainers and from student to student. Although all experts met and reviewed videos before the trial in a genuine attempt to ensure consistency of the feedback they would provide, there will be a natural variability between assessors. However, this also represents what would occur in real life, as opposed to having standardised videos that students could use as a resource for learning. Genuine consistency across teaching and assessment is challenging to accomplish. Every attempt was made to recruit enthusiastic clinicians with a genuine interest in medical education, as well as some of the aspects that may be attributed to good educators.[40]

More research needs to be carried out to determine what impact UVF can have on learning simple and more advanced clinical and surgical skills. It would also be interesting to determine what impact both DEF and UVF could have when used in combination. Students may gain additional benefit from being able to learn the basics of a skill with UVF, and then have an expert hone what they have learned. Alternatively, it may prove to be more beneficial to have an expert teach the skill at an early stage and use UVF to practice what has been taught to maintain and further perfect a skill.

The results of this study imply that performance of simple clinical skills can be enhanced by the ability to self-review performances. The need for faculty has been previously suggested as a reason why simulation training is not integrated into training programmes.[41] These results suggest that the presence of faculty confers no advantage over self-review when basic practical skills are being taught. UVF, with quality-ensured training videos, could potentially be a resource-efficient method of teaching and improving clinical skills.

REFERENCES

1. Mahmood T, Darzi A. The learning curve for a colonoscopy simulator in the absence of any feedback: no feedback, no learning. *Surg Endosc*. 2004;18(8):1224-1230.
2. National Student Survey: Findings and Trends 2006-2010. ⟨www.hefce.ac.uk/media/hefce/content/pubs/2011/201111/11_11.pdf⟩ Accessed 24.05.12.
3. Johnson G, Barrett J, Jones M, Parry D, Wade W. Feedback from educational supervisors and trainees on the implementation of curricula and the assessment system for core medical training. *Clin Med*. 2008;8(5):484-489, http://www.ncbi.nlm.nih.gov/pubmed/18975478. Accessed 21.03.16.
4. Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: are they just a tick box exercise? *Med Teach*. 2011;33(11):919-927.
5. Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. *N Engl J Med*. 2006;355(25):2664-2669.
6. Beard JD, Marriott J, Purdie H, Crossley J. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technol Assess*. 2011;15(1):i-xxi,1-162.
7. Williams CT, Fost N. Ethical considerations surrounding first time procedures: a study and analysis of patient attitudes toward spinal taps by students. *Kennedy Inst Ethics J*. 1992;2(3):217-231, http://www.ncbi.nlm.nih.gov/pubmed/10121088. Accessed March 20.03.16.
8. Graber MA, Wyatt C, Kasparek L, Xu Y. Does simulator training for medical students change patient opinions and attitudes toward medical student procedures in the emergency department? *Acad Emerg Med*. 2005;12(7):635-639.
9. Graber MA, Pierre J, Charlton M. Patient opinions and attitudes toward medical student procedures in the emergency department. *Acad Emerg Med*. 2003;10(12):1329-1333, http://www.ncbi.nlm.nih.gov/pubmed/14644784. Accessed 20.03.16.

10. Farquharson AL, Cresswell AC, Beard JD, Chan P. Randomized trial of the effect of video feedback on the acquisition of surgical skills. *Br J Surg*. 2013;100(11):1448-1453.

11. Brydges R, Carnahan H, Safir O, Dubrowski A. How effective is self-guided learning of clinical technical skills? It's all about process *Med Educ*. 2009;43(6):507-515.

12. Brydges R, Nair P, Ma I, Shanks D, Hatala R. Directed self-regulated learning versus instructor-regulated learning in simulation training. *Med Educ*. 2012;46(7):648-656.

13. Jensen AR, Wright AS, Levy AE, et al. Acquiring basic surgical skills: is a faculty mentor really needed? *Am J Surg*. 2009;197(1):82-88.

14. Jowett N, LeBlanc V, Xeroulis G, MacRae H, Dubrowski A. Surgical skill acquisition with self-directed practice using computer-based video training. *Am J Surg*. 2007;193(2):237-242.

15. Brydges R, Dubrowski A, Regehr G. A new concept of unsupervised learning: directed self-guided learning in the health professions. *Acad Med*. 2010;85(10 Suppl):s49-s55.

16. Nesbitt CI, Phillips AW, Searle RF, Stansby G. Randomized trial to assess the effect of supervised and unsupervised video feedback on teaching practical skills. *J Surg Educ*. 2015;72(4):697-703.

17. General Medical Council. *Tomorrow's Doctors: Duties of a Doctor Registered With the General Medical Council*. London; 2009.

18. Greveson GC, Spencer JA. Self-directed learning—the importance of concepts and contexts. *Med Educ*. 2005;39(4):348-349.

19. General Medical Council. *Tomorrow's Doctors: Outcomes and Standards for Undergraduate Medical Education*. ⟨http://www.gmc-uk.org/Tomorrow_s_Doctors_1214.pdf_48905759.pdf⟩ Accessed 18.03.16.

20. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med*. 1999;74(10):1129-1134, http://www.ncbi.nlm.nih.gov/pubmed/10536636. Accessed 16.02.16.

21. Hodges B, McNaughton N, Regehr G, Tiberius R, Hanson M. The challenge of creating new OSCE measures to capture the characteristics of expertise. *Med Educ*. 2002;36(8):742-748, http://www.ncbi.nlm.nih.gov/pubmed/12191057. Accessed 20.03.16.

22. Govaerts MJB, van der Vleuten CPM, Schuwirth LWT. Optimising the reproducibility of a performance-based assessment test in midwifery education. *Adv Health Sci Educ Theory Pract*. 2002;7(2):133-145, http://www.ncbi.nlm.nih.gov/pubmed/12075145. Accessed 20.03.16.

23. Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. *Acad Med*. 2003;78(2):219-223, http://www.ncbi.nlm.nih.gov/pubmed/12584104. Accessed 20.03.16.

24. Shippey S, Handa VL, Chen TL, Chous B, Bowen CW. Validation of an instrument for evaluation of subcuticular suturing using a plastic tissue model. *J Surg Educ*. 2009;66(1):31-34.

25. Acton RD. The evolving role of simulation in teaching surgery in undergraduate medical education. *Surg Clin North Am*. 2015;95(4):739-750.

26. Gardner AK, Scott DJ, Pedowitz RA, et al. Best practices across surgical specialties relating to simulation-based training. *Surgery*. 2015;158(5):1395-1402.

27. Cachia M, Pace-Bardon M, Balzan G, Tileny R, Micallef J, Balzan M. Simulation training for foundation doctors on the management of the acutely ill patient. *Adv Med Educ Pract*. 2015;6:657-663.

28. Caliendo E, Kopacz R. *Improving Student Motivation and Performance in Music Programms*; 1999.

29. Christina R, Barresi J, Shaffner P. The development of response selection accuracy in football linebacker using video training. *Sport Psychol*. 1990;4(1):11-17.

30. Winfrey M, Weeks D. Effects of self-modelling on self efficacy and balance beam performance. *Appl Res Coach Athl Ann*. 1996;14:523-543.

31. Lammers RL, Temple KJ, Wagner MJ, Ray D. Competence of new emergency medicine residents in the performance of lumbar punctures. *Acad Emerg Med*. 2005;12(7):622-628.

32. Schmidt RA, Wulf G. Continuous concurrent feedback degrades skill learning: implications for training and simulation. *Hum Factors*. 1997;39(4):509-525.

33. Schmidt R a, Bjork R. New conceptualisations of practice—common principles in three paradigms suggest new concepts for training. *Pschol Sci*. 1992;3(4):207-217.

34. Salmoni AW, Schmidt RA, Walter CB. Knowledge of results and motor learning: a review and critical reappraisal. *Psychol Bull*. 1984;95(3):355-386.

35. Singh P, Aggarwal R, Tahir M, Pucher PH, Darzi A. A randomized controlled study to evaluate the role of video-based coaching in training laparoscopic skills. *Ann Surg*. 2015;261(5):862-869.

36. Zevin B, Aggarwal R, Grantcharov TP. Surgical simulation in 2013: why is it still not the standard in surgical training? *J Am Coll Surg*. 2014;218(2):294-301.

37. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004;135(1):21-27.

38. Strandbygaard J, Bjerrum F, Maagaard M, et al. Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator. *Ann Surg*. 2013;257(5):839-844.

39. Zendejas B, Brydges R, Hamstra SJ, Cook D. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Ann Surg*. 2013;257(4):586-593.

40. Harden R, Crosby J. AMEE Guide No 20: the good teacher is more than a lecturer—the twelve roles of the teacher. *Med Teach*. 2000;34:334-337.

41. Okuda Y, Bond W, Bonfante G, et al. National growth in simulation training within emergency medicine residency programs, 2003-2008. *Acad Emerg Med*. 2008;15(11):1113-1116.

This was a follow up study, built on the initial investigation by expanding the number of students in each group and adding two additional skills. A generic lecture feedback group was not used and thus the aim was to establish whether a difference existed between what could be considered the gold standard of feedback (expert on a one-to-one basis) versus students reviewing their own performance with an expert video as a comparator. Again the results established that students' self-evaluations were not inferior to having an expert provide feedback. Like the pilot study the skills involved in the follow-up trial were basic, but the added number of candidates included helps to build on the findings of the original study.

There are perhaps two major questions that arise as a consequence of both of these studies. The first relates to what is the "learning" significance of these statistically significant findings. Whilst a significant level of improvement was demonstrated both with respect to checklist scores and global skills, it is difficult to determine whether statistical findings equates to students moving from being unable to perform a task safely to becoming more competent. However, it would be reasonable to expect that having the technology to record oneself performing a task and being able to self-criticise would aid learning a skill.

The second question regards *learning* a skill as opposed to be performing a skill. The findings from this study demonstrate that there is an improvement in performance over a short time period. Learning would equate to skill retention over a prolonged period of time. This requires further investigation but the facility to review one's own performance is likely to enhance the learning experience. This mechanism of learning should help reduce the burden on trainer time and develop the skills for self-appraisal. There is a known performance-learning paradox, where short-term improvement does not equate to long-term learning. It may be that the unsupervised feedback used within this study engages student self-discrimination and allows skills to be learned.

3.3.3 Use of domain versus tick box assessment

Within both of these studies checklist scores and global scores were used to score candidates. The use of both of these measures is common place within OSCEs and other high stakes measures as a means of standard setting. This involves plotting checklist scores against global scores and using regression analysis to determine the pass mark. Its benefit is in assessments where the examiner is able to witness the performance of the candidate and

formulate an impression of their ability as well as scoring on a checklist. The benefits of this method include its ease of use for examiners, and reliability. It may also be used to identify problems with a particular assessment station. It is however vital that examiners understand what constitutes a borderline candidate so that the global impression formed is consistent.

In these manuscripts no pass or fail mark was determined. Rather the aim was to determine whether an improvement in scores can be achieved using the feedback mechanisms employed and whether global scores saw similar improvement.

3.4 Discussion

One of the advantages of the more technologically advanced simulators is their ability to give feedback. This may be on the efficiency of movement to complete a task, or a record of any mistakes made. Feedback is vital component in learning surgical skills, often providing a source of motivation and information[75]. Having this as an inbuilt tool greatly enhances the simulator as a device for learning. It is also important that an individual reflects on their own performance when using these tools. Resuscitation models have developed to the point that they can have audible heartbeats and heart traces that may change as different interventions are instituted. One of the most realistic features of these models is that even after doing everything correctly the "patient" may still die. The value of feedback from the trainer and the simulator (where possible) is invaluable. Unfortunately, in real life, even when everything is done correctly the desired outcome is not always attained.

Mastery of a surgical skill is achieved through drill, repetition and practice[79]. Simulators proved an ideal aide to this, in a safe reproducible environment. Skills can be learnt individually and gradually built up until all the steps can be put together. Further practice can be continued until a level of automaticity is reached. In a well-designed simulator these skills should be directly transferable to real life scenarios. Skills learnt using simulators have been shown to be directly transferable to real-life situations[80,81]. However, nothing is quite like performing the procedure on a real patient. While confidence that one can complete the task should have been gained by using these tools there is always a higher level of stress involved in doing it in "real life" the first few times.

With the training hours available for surgeons on the decline, reliance on volume of exposure to achieve the appropriate skills will soon not be possible. In this country compliance with the

EWTD, the 48-hour-week and no extension to the years spent in training mean that an alternative method to learning and honing skills is required. Simulators will allow training to be tailored to the student's needs, and will allow students to be able to practice the skills they feel weak at until techniques become second nature.

3.5 Chapter Conclusions

Whilst WBAs have become widely used and accepted within the surgical curriculum there remains some concern regarding their overall reliability and validity. They have the potential to enhance training and ensure timely and useful feedback is provided. A lack of quality feedback is a common complaint from students both medical and from outside medicine[82]. One of the important benefits of performing WBAs is as an aid to picking up trainees that are struggling. Their proper use may help identify those in difficulty at an early stage and serve as an aid to remedying problems.

Whilst WBAs are a key feature of ensuring feedback within the new curriculum, changes in training have meant an increasing trend to use towards technology and simulation to aid training and provide feedback. The two randomised trials included within this thesis have determined that review of one's performance, either with an expert, or by self-evaluating can lead to a significant improvement in performance of a simple clinical skill. Whilst it is not clear if the findings from these papers can be extrapolated to more advanced technical skills, they may ease the burden on requiring trainers for simpler skills, and could be used to help teach more complicated clinical skills in stages. Further evaluation into the use of this technology is required to establish how it can be best used.

Chapter 4: Perceptions of Formative Assessments.

Theme 3: Perceptions of formative assessment

1. Phillips AW, Madhavan A, Bookless LR, Macafee DA. Surgical Trainers' Experience and Perspectives on Workplace Based Assessments. *J Surg Educ*. 2015 Sep-Oct;72(5):979-84[6]
2. Phillips A**,** Lim J, Madhavan A, Macafee D. Case-based discussions: UK surgical trainees' perceptions. *Clin Teach*. 2016 13(3): 207-212[7]
3. Nesbitt CI, Phillips AW, Searle RF, Stansby G. Student views on the use of two styles of video-enhanced feedback compared to standard lecture feedback during clinical skills training. *J Surg Educ*. 2015 Sep-Oct;72(5):969-73[8]

This chapter looks at student perceptions of the feedback that has been discussed through the previous chapters. The first paper reviews the perceptions of trainers on WBAs. There has been little work looking at their views (as opposed to trainees) and the aim was to try and identify what issues consultant trainers had.

The second paper involves a qualitative study on the use of CBDs in Surgical Training. This was a large study producing important themes that trainees raised regarding CBD usage that are applicable to all WBAs. Given the restrictions of the journal article for discussing this piece of work an expanded discussion of the methodology and limitations has been included.

The final paper looks at student views on the video feedback that was provided for the trials in the previous chapter. This was to determine whether the self-review was regarded as acceptable and beneficial.

4.1 Introduction

Training programmes, both undergraduate and postgraduate, aim to produce clinicians that are able to deliver a high level of patient care. In order to do this training must be of a high quality. One large retrospective study, involving over five million cases performed by over 4000 surgeons and evaluating outcomes from US surgical residency programmes demonstrated that graduates from the top quintile of a training programme experienced 30% fewer complications than those from the lowest quintile[83].

Chapter 2 illustrated how surgical training has shifted in the UK, and has shown that good quality training exists, and technically demanding procedures can be learned within the confines of the new curriculum. Chapter 3 has looked at the validity and reliability of WBAs which have been integrated into the new curriculum, and evaluated the use of other feedback mechanisms in learning simple technical skills.

The aim of the papers in this chapter was to evaluate the perspectives of trainers and trainees on these mechanisms for enhancing training.

4.2 Why Study Perceptions?

Understanding the perceptions of those involved in using WBAs, both as students and as trainers, is vital in determining how they are used day-to-day. Each assessment can be thought of as a discrete encounter, the parameters of which are established by the trainer and trainee and may vary between encounters, and as those involved change. In essence, no two WBAs will be the same, but how one trainer conducts a CBD or PBA may be very different from how another conducts it.

Views regarding the actual mechanism of the assessment may influence how the process is enacted. A number of other parameters may influence how it is conducted including the relationship between assessor and trainee, the perceived difficulty of the task and the understanding of the stakes involved. Although guidelines exist as to how WBAs should be conducted, and many trainers will have undergone courses, WBAs may be adapted to suit the purposes of the individuals involved. Thus trainees' and trainers' perceptions of how they participate in and enact WBAs are important, in terms of both how assessments are constructed and how they are utilised.

4.3 Surgical Trainers' Experience and Perspectives on Workplace Based Assessments. Paper and commentary

Phillips AW. Madhavan A, Bookless LR, Macafee DA. Surgical Trainers' Experience and Perspectives on Workplace Based Assessments. *J Surg Educ*. 2015 Sep-Oct;72(5):979-84[6]

Surgical Trainers' Experience and Perspectives on Workplace-Based Assessments

Alexander W. Phillips, MA, FRCSEd (Gen Surg), MFSTEd, Anantha Madhavan, MBChB, MRCS,† Lucy R. Bookless, MBChB, MRCS,* and David A.L. Macafee, MA, DM, FRCS†*

*Department of General Surgery, Royal Victoria Infirmary, Newcastle upon Tyne, United Kingdom; and †Department of General Surgery, James Cook University Hospital, Middlesbrough, United Kingdom

BACKGROUND: Workplace-based assessments (WBAs) were designed to provide formative feedback to trainees throughout their surgical career. Several studies highlight dissatisfaction with WBAs, and some feel they lack validity and reliability and exist as a "tick-box exercise." No studies have looked at the attitudes of the assessor.

AIM: The aim of this study was to evaluate perceptions and experience of the 4 intercollegiate surgical curriculum programme WBAs by assessors.

METHODS: An 18-item electronic questionnaire, including 6-point Likert scoring questions, was e-mailed to all surgical program directors for distribution to general surgery trainers within their deanery.

RESULTS: In total, 64 responses were received. All trainers had been trained in using WBAs. Trainers had the most experience with procedure-based assessments (PBAs)—72% of trainers had completed more than 25 PBAs. Trainers felt PBAs were the most beneficial WBA, and both PBAs and case-based discussions were regarded as significantly more useful than mini–clinical evaluation exercise ($p < 0.05$). More than 74% stated that WBAs were mainly initiated by trainees, and only 10% had specific sessions allocated to complete WBAs.

CONCLUSION: WBAs are regarded as beneficial to trainees. The results suggest that assessors feel case-based discussions and PBAs, which assess higher thinking and practice of complex practical skills, respectively, are significantly more useful than assessments involved in observing more straightforward clinical and procedural interactions. (J Surg Ed 72:979-984. © 2015 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: workplace based assessments, formative assessment, intercollegiate, case based discussion, procedure based assessment, direct observation of procedural skills, mini clinical evaluation exercise

COMPETENCIES: Practice-Based Learning and Improvement

INTRODUCTION

Workplace-based assessments (WBAs) have been adopted throughout medical curricula. Within the United Kingdom, they have become an integral component of the new surgical curriculum—the intercollegiate surgical curriculum programme (ISCP).[1] Their primary function is to provide formative feedback to surgical trainees through a variety of styles. These included providing feedback for practical skills using direct observation of procedural skills and procedure-based assessments (PBAs) and enhancing clinical learning using mini–clinical evaluation exercise (mini-CEX) and case-based discussions (CBDs). Furthermore, they serve to provide evidence of learning, attainment of skills, and aid trainers in identifying those struggling in performing tasks. The value of WBAs has been strongly linked to trainer implementation[2]; thus, determining assessor perceptions is an important factor in potentially improving these assessments' potential effect.

Current surgical training in the United Kingdom is divided into early "core training" years, where trainees are required to achieve relevant competencies and the membership exam, before participating in a national selection process to gain entry into "higher surgical training." At each of these levels, trainees rotate between posts that usually last 6 or 12 months. The entire program is held together by the ISCP. This was introduced in 2007 and serves to act as a complete curriculum that uses an online

Correspondence: Inquiries to Alexander W. Phillips, MA, FRCSEd (Gen Surg), MFSTEd, Department of General Surgery, Royal Victoria Infirmary, Newcastle upon Tyne, UK; e-mail: awphillips@doctors.net.uk

TABLE 1. Demographics of the Assessors

Male:female	54:10
Full time:part time	60:4
Number of years qualified as consultant	
<5	13 (20%)
5-10	24 (38%)
>10	27 (42%)
Used WBAs as a trainee	11 (17%)

platform enabling trainees to maintain a portfolio, access the syllabus, and record formative assessments. As it has evolved, areas for recording other achievements and reflections have been added. Trainees are required to demonstrate satisfactory progression at annual reviews of competence progression and use WBAs as evidence toward this.[3]

Several studies have highlighted a level of dissatisfaction with the WBAs that form part of this new curriculum. Trainees have commented that they felt they lacked validity and reliability and have become no more than a tick-box exercise.[4] A part of this stems from a lack of engagement with these new tools by both trainers and those being assessed. The ISCP curriculum placed an emphasis on trainees to take responsibility for their own training and ensure that they had obtained the requisite assessments to evidence their competence. In addition, deaneries produced their own guidance on how many WBAs of each type are required in each year of training. It is emphasized that trainees regularly perform WBAs throughout posts rather than completing them in the last few weeks where they can be more confident in performing well.[5]

WBAs have the ability to help demonstrate both competence and performance. The former forms the penultimate level of Miller's pyramid (shows how)—being able to meet a preset checklist of what is required—whereas the latter demonstrates ability at the highest level of Miller's pyramid (does) and is close to representing actual practice.[6] These assessments may serve to formalize training techniques that have been occurring for generations; however, their real value, provided they are used correctly, is ensuring trainees get appropriate and regular feedback and as such should be used regularly throughout each training post.

Despite the debate about their usefulness, and several studies looking at the attitudes of trainees, no studies have looked at the attitudes of those participating in these assessments as the assessor in any medical specialty.

The aim of this study was to survey surgical trainers in the United Kingdom who are frequently assessing WBAs to evaluate their perceptions and experience of how these tools are being employed within the ISCP.

METHODS

A web-based electronic survey was undertaken. This comprised 18 questions, including demographics, the experience of assessors with WBAs, their training, and perceptions and usage of WBAs. Likert scoring questions were based on a 6-point scale to discourage respondents from taking neutral viewpoints. Furthermore, space for free-text comments on opinions of WBAs was provided. The questionnaire was designed along the principles by Cohen et al.[7]

General surgery deanery program directors were contacted by e-mail and were asked to forward the questionnaire to consultant general surgical trainers within their deanery. Further request e-mails were sent on 2 occasions to those deaneries where no responses were obtained at 1-month intervals.

Data Analysis

Data from all questionnaires was collated, and statistical analysis with a Mann-Whitney U test was used for comparison of Likert scoring questions. A $p < 0.05$ was considered significant.

Thematic analysis was performed manually. An initial coding structure was developed and evolved with iteration to identify the main themes that emerged.

RESULTS

In total, 18 program directors were contacted. A total of 64 responses from general surgeons were received from 9 different deaneries. Of those that responded, 54 (85%) were men and 60 (95%) were employed full time (Table 1). Only 11 (17%) had experience of using WBAs as a trainee. Figure 1 shows the type of training assessors have been given

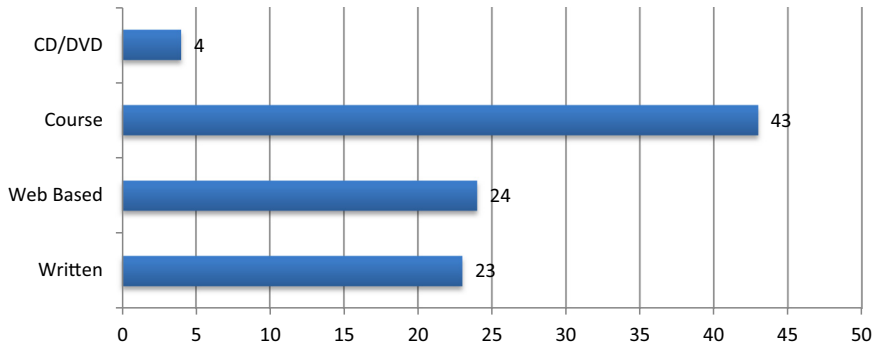

FIGURE 1. How trainers have been educated as to workplace-based assessment use.

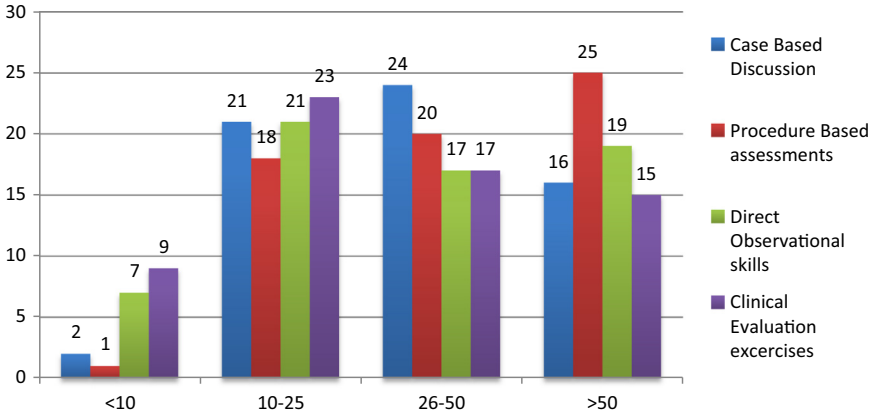

FIGURE 2. The trainers' experience in work-based assessments.

for using WBAs; all trainers had received at least 1 form of instruction in their usage. A mean score of 5 of 6 was reported on the question of "how knowledgeable are you on how WBAs should be conducted."

Assessors indicated varying experience with each of the 4 types of WBA being evaluated. PBAs had clearly been employed most frequently by trainers, with the mini-CEX being used least frequently (Fig. 2).

It was noted that few trainers in the survey indicated that they initiated WBAs. More than 74% of trainers stated that the WBAs were nearly always initiated by the trainees. Furthermore, only 10% of trainers stated they had specific sessions allocated for conducting WBAs.

Trainers had varied opinions on WBAs. More than 65% of trainers felt that WBAs aided higher surgical trainees (HSTs). However, only 30% of trainers felt WBAs were useful for training core surgical trainees (CSTs). Mini-CEXs were deemed a more useful tool for assessments in CSTs than HSTs, although this did not reach statistical significance. Further, 70% of trainers felt that PBAs were the most beneficial WBA in training (Table 2). Overall, 80% felt trainees need to complete a minimum of 15 PBAs per year. In comparison with PBAs, trainers felt fewer CBDs and CEXs needed to be completed annually, with 60% expressing trainees should only need to complete a minimum of 10 or less a year (Fig. 3). PBAs and CBDs were regarded as being significantly more useful than mini-CEXs ($p < 0.05$), and PBAs were seen as significantly more useful than direct observation of procedural skills (Tables 3 and 4).

Slightly more than half (56%) of trainers felt the performance of trainees at WBAs had an effect on altering a trainees learning needs. Consensus among trainers was that minimum number of assessments required per year of training should be varied depending on the individual and their level of training.

Free-text comments revealed a number of recurring themes from the assessors (Table 5). The most frequent comments pertained to a lack of time to perform these assessments; this appeared to be largely owing to no free time in the working week, the feeling that they were time consuming to complete, and that there was sometimes a "rush" toward the end of a placement to fill the requisite numbers of assessments. There was also a feeling that assessments should not be chosen by the trainees as this allows them to pick cases they are comfortable with. Other commonly expressed views included that they were only useful if done properly. Some felt they had not received adequate training and that the online form did not allow any flexibility. Finally, a number of trainers expressed the view that WBAs added to an ever-increasing burden of "paperwork" (Table 5).

DISCUSSION

WBAs have become an integral component of most medical curricula, and surgical trainees in the United Kingdom must undertake a mandatory number of these assessments each year.[8] Several studies have reviewed their use and the attitudes of trainees, but hitherto no studies have evaluated the attitudes of those involved in doing the assessment.[8-10]

TABLE 2. Likert Scale Illustrating Trainers Views of the Usefulness of WBAs as an Educational Tool

Work-Based Assessment	1 (Not Valuable)	2	3	4	5	6 (Very Valuable)	Mean Rating
Case-based discussions	2	9	10	12	23	8	4.08
Clinical evaluation exercises	4	10	18	11	17	4	3.59
Direct observational skills	4	7	12	11	23	7	3.93
Procedure-based assessments	2	6	9	6	22	19	4.51

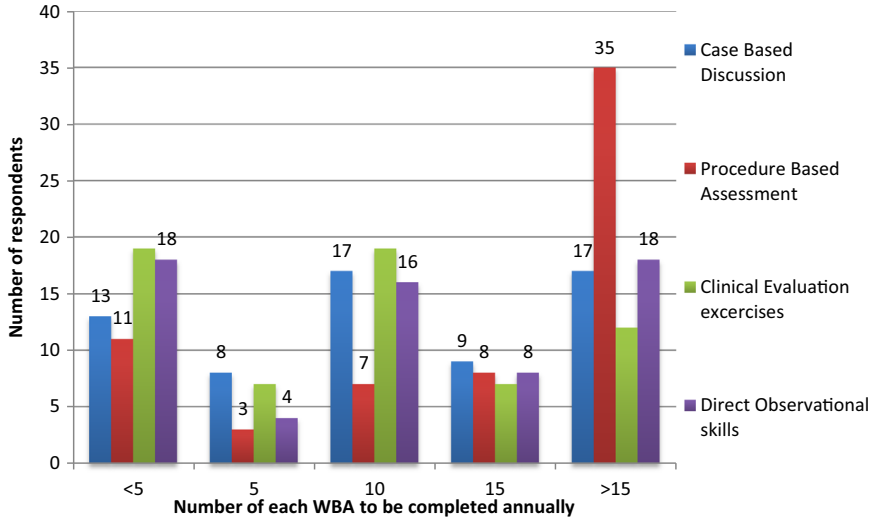

FIGURE 3. Respondents' views on how many of each type of WBA trainees should complete.

The results from this study have demonstrated that generally these assessments were perceived as beneficial to trainees, with PBAs being particularly regarded in a positive light. These are nearly always used with HSTs to assess more advanced technical tasks and allow an operation to be broken down into steps, enabling complicated procedures to be learnt in stages and also for improvement to be demonstrated. CBDs were also viewed as very beneficial, with trainers feeling they were equally beneficial to HSTs and CSTs. The advantage of the CBD is that it can be tailored to a trainee's level and be used to encourage higher thinking. Mini-CEXs were viewed as less useful and perceived as being less beneficial to HSTs than to CSTs, although this did not reach statistical significance. This might be a reflection on the fact that HSTs should be competent at many of the tasks mini-CEXs are good for assessing, including history taking, examination, and gaining consent. Further, trainers had less experience at using mini-CEXs possibly because they feel that it is a less useful tool and consequently only bare minimum numbers of these are completed; another explanation may be that the mini-CEX is frequently conducted by nonconsultant assessors—such as HSTs for core trainees, resulting in consultants having less experience with this assessment. In contrast, trainers had more experience with PBAs and CBDs, reflecting the additional value placed on these assessments.

Despite their intended use as formative assessment tools, there is increasing evidence that WBAs are being incorrectly viewed by trainers and trainees as a summative, rather than a formative, assessment.[8] There appears to be a tendency for the current era of postgraduates, traditionally subjected to summative examinations at every stage of their lives, to approach WBAs with a similar, summative mindset, thereby minimizing their underlying objective. This may be exacerbated by the small number of consultants who have actually used these assessment tools as the trainee, and then during their inception and piloting. The General Medical Council feel that WBAs should be used as formative assessments, and there has been a move to rebrand these "assessments" across all specialties as supervised learning events to emphasize their importance in learning rather than assessment.[11] (supervised learning event should be regarded as an umbrella term that includes any interaction between trainer and trainee that leads to feedback and learning). With the appointment of new consultants who have had experience with WBAs within their own training, their formative importance may be more fully appreciated.

The free-text comments revealed that many trainers felt they had inadequate time to perform assessments and that

TABLE 3. Comparison of Likert Scores for Each WBA

Compared WBAs	p Value
PBA vs CBD	0.0512
PBA vs mini-CEX	**0.0003**
PBA vs DOPS	**0.0193**
CBD vs mini-CEX	**0.048**
CBD vs DOPS	0.61
Mini-CEX vs DOPS	0.16

DOPS, direct observation of procedural skills.
Bolded p-values are significant.

TABLE 4. Likert Scale Illustrating the Trainers' Views on the Usefulness of WBAs for Core Surgical Trainees (CSTs) and Higher Surgical Trainees (HSTs)

Work-Based Assessment	CST Mean Likert score	HST Mean Likert score	p Values
Case-based discussions	4.03	3.97	0.94
Clinical evaluation exercises	3.84	3.59	0.41

TABLE 5. Thematic Breakdown and Representative Comments From Free-Text Section Asking Trainers for Their Opinions on WBAs

Assessors Free-Text Comments	Theme
More time is needed to use them properly.	Time to do assessments
Time in job plan very inadequate.	
Very little time in a consultant's schedule to do justice to everything about ARCPs. Trainees invariably bring these by the end of the rotation.	
Perception that they take a long time is off-putting for trainers and trainees, but they can easily be incorporated into routine practice.	
A good idea—short on quality and evidence, I'm afraid.	Value of WBAs
I have been trained in the use of WBAs and understand the theory and practice of them. I have to say that I think their greatest use is as an educational tool either to cover areas where a trainee has been lacking or to provide evidence that certain areas of the curriculum have been covered if this cannot be done through normal clinical practice.	
Valuable if utilized appropriately.	
I don't have any experience of how the data generated by WBAs is used—e.g., by educational supervisors, or at ARCPs.	
Need to be formalised, i.e., the trainee can be selective about the cases that they wish to submit for assessments. There should be no choice–all should be included; the results should be formally assessed to ensure progress against a standard.	Case selection
I think it has added very little to training except more paperwork. We performed these before but didn't necessarily document them.	Increased "paperwork"
Sometimes the WBAs are a bit nonflexible to use in certain situations.	Flexibility of assessments

ARCP, annual reviews of competence progression.

this perhaps needs to be given more emphasis in job planning. The time-consuming nature of WBAs has been previously documented as reasons for them not being performed.[12,13] There have been some suggestions that they should be written into the job plan of trainers to ensure they are conducted[13] or planning them into clinical activities to try and minimize any financial burden.[14] It was suggested by some respondents that WBAs have indeed been performed through the ages but ISCP serves to formalize "what we have been doing all along." Although this may be true, formalizing these assessments confers a number of benefits. It allows both trainees and trainers to monitor progress over time and should give trainees more regular one-to-one interaction with their trainer. Several trainers commented that trainees should not be choosing cases as it allows them to be selective, which may potentially be not challenging and not beneficial; paradoxically, very few allocated specific sessions in their timetable to conduct these assessments, or admitted to initiating WBAs, which would potentially mean WBAs were more challenging and were more useful. Formalizing these WBAs also encourages "triangulation" of assessment, with multiple assessors conducting the same assessment and on multiple occasions, which is needed for each assessment to be reliable and valid.

The electronic portfolio structure of the ISCP has gone a long way in trying to minimize the "added paperwork" perception of these assessments. However, some hospital technology may make completing them painfully slow, and there has been suggestion that access to computers can also be inhibiting.[15] Phone "apps" have been developed that may help speed up the process; moreover, the ISCP now has its own app to facilitate easy WBA completion. A further option would be to encourage trainees to conduct regular self-assessment using the same WBAs, which will ensure they think and reflect on their daily care, and these can then be reviewed by the educational supervisor who can correlate self-assessments with their own interactions with the trainee. This may serve to reduce the amount of "paperwork" for assessors and establish a greater portfolio of evidence.

The trainers who participated in this study were a representation of those involved in training from across the country. Invitations were sent through the program director in each deanery. In total, 64 responses were obtained, which is likely to be large enough to have obtained responses from those with a spectrum of attitudes toward WBAs. However, the authors estimate that this represents less than 10% of those involved in training, which is a potential shortcoming of this study. Furthermore, this type of survey is subject to self-selection bias in that responses are more likely to come from those who are enthusiastic about the use of WBAs; however, equally it may lead to those who have strong feelings against these tools participating. Despite this, the demographics suggested that the views were not greatly polarized, and participants had good experience using WBAs. All trainers who responded had received some form of training in WBAs, which is a much higher proportion than that was found among psychiatry trainers,[16] and most trainers felt that they had a good knowledge of how work-based assessments should be conducted. However, it is difficult to determine whether the assessors' perceptions of how they should be conducted correlates with ISCP intentions. Indeed this requires further evaluation. Both within this study and in the studies that have investigated trainee

viewpoints, it was felt that any failing of a WBA was often due to incorrect use.[9] A previous study reviewed the effect of assessor training on performance and found that there was little difference in raters who had attended a workshop and their control group who had not, although they did find that the course improved trainers confidence in using the tool.[17] There has been a suggestion that clinical rater training programs are of little benefit,[18] possibly owing to the programs themselves being ineffectual, or possibly due to raters being "impervious to training."[19] There has been no formal review to investigate the effect of various training methodologies on how surgical trainers conduct WBAs, but the mixed comments from those in this study suggest a split between those that have embraced conducting the assessments and those that feel they are of limited educational value, a group that may be impervious to training.

A review of WBAs by the Academy of Medical Royal Colleges of the United Kingdom in 2009 suggested widespread confusion regarding standards, methods, and goals of WBAs, which has prompted increased cynicism within the profession.[20] Our results suggest that assessors feel there is a real benefit for trainees in conducting WBAs particularly when used to teach complex practical skills (PBAs) and when used to encourage higher thinking (CBDs), whereas trainers feel they need more time to perform the assessments in the correct fashion and that the online forms provided need to allow a greater degree of flexibility.

REFERENCES

1. *Intercollegiate Surgical Curriculum Programme*. Available at: ⟨www.iscp.ac.uk⟩; 2011.
2. Nesbitt A, Baird F, Canning B, Griffin A, Sturrock A. Student perception of workplace-based assessment. *Clin Teach*. 2013;10(6):399-404.
3. Phillips AW, Madhavan A. A critical evaluation of the intercollegiate surgical curriculum and comparison with its predecessor the "Calman" curriculum. *J Surg Educ*. 2013;70(5):557-562.
4. Ali JM. Getting lost in translation? Workplace based assessments in surgical training. *Surgeon*. 2013;11(5):286-289.
5. *Academy of the Royal Medical Royal Colleges*. Workplace based assessment forum outcomes. Available at: ⟨http://www.aomrc.org.uk/doc_view/9320-workplace-based-assessment-forum-outcomes-2010⟩; 2010.
6. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(suppl 9):S63-S67.
7. Cohen L, Manion L, Morrison K. Research Methods in Education. 5th ed. London: Routledge Falmer; 2000.
8. Pereira EAC, Dean BJF. British surgeons' experiences of a mandatory online workplace based assessment portfolio resurveyed three years on. *J Surg Educ*. 2012;70(1):59-67.
9. Pereira EAC, Dean BJF. British surgeons' experiences of mandatory online workplace-based assessment. *J R Soc Med*. 2009;102(7):287-293.
10. Hunter AR, Baird EJ, Reed MR. Procedure-based assessments in trauma and orthopaedic training—The trainees' perspective. *Med Teach*. 2014;4:1-6 [Epub ahead of print].
11. *GMC Education Discussion Group*. Learning and assessment in the clinical environment: the way forward—November 2011; 2011.
12. Wilkinson JR, Crossley JGM, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ*. 2008;42(4):364-373.
13. Morton j, Cumming A, Cameron H. Performance-based assessment in undergraduate medical education. *Clin Teach*. 2007;4(1):36-41.
14. Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: are they just a tick box exercise? *Med Teach*. 2011;33(11):919-927.
15. Menon S, Winston M, Sullivan G. Workplace-based assessment: survey of psychiatric trainees in Wales. *Psychiatry Bull*. 2009;33(12):468-474.
16. Babu K, Htike M, Cleake V. Workplace based assessments in Wessex: the first 6 months. *Psychiatry Bull*. 2009;33(12):474-478.
17. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med*. 2009;24(1):74-79.
18. Woehr D, Huffcutt A. Rater training for performance appraisal: a quantitative review. *J Occup Organ Psychol*. 1994;67(3):189-205.
19. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med*. 2003;15(4):270-292.
20. Collett A. Improving Assessment. Academy of the Medical Royal Colleges; 2009.

A number of studies have evaluated what trainees have thought about WBAs. Pereira and Dean evaluated student views of the new ISCP curriculum at two years after its inception and then four years after that[84,85]. Their initial findings indicated a great deal of dissatisfaction with ISCP and a feeling that there was a lack of validity with the formative assessments that were integrated into it. Their online survey of over 500 surgical trainees indicated 49% of trainees felt the assessments were poor or very poor, and 41% felt that the changes had impacted negatively on their training. Their follow up survey four years later demonstrated that trainees generally felt there had been some improvement in with ISCP with much less negative feedback and an overall feeling that the new curriculum had a "neutral" impact on training.

Other studies have shown that medical students have negative perceptions towards WBAs[86]. Ali and Goh received responses from 115 final year medical students who indicated that 90% had benefited from WBAs[86]. However, perception of WBAs was influenced by whether they had a good understanding of WBAs. A number of students had negative perceptions towards WBAs which could potentially continue on into their postgraduate training. De Jonge et al found that different perspectives affect acceptance of WBAs[87]. Interestingly students identified a number of obstacles to WBA use that correlate well with the obstacles that postgraduate students and trainers identified[6,7]. These obstacles included a lack of time, lack of trainer availability, lack of engagement and a view that WBAs are essentially a tick box exercise.

This survey on trainers' experience was the first attempt to evaluate what trainers rather than trainees thought about WBAs. It was carried out as an online questionnaire circulated to all the surgical programme directors in the UK who were asked to forward the online questionnaire to all the educational supervisors in General Surgery within their Deanery. It is difficult to ascertain what the response rate actually was, and only 64 replies were obtained which is likely to be under 10% of the educational trainers nationwide. However, it was felt that this did provide a snapshot of trainers' perspectives.

The findings from this study were interesting, particular as they provided a first impression of what the *trainers* rather than *trainees* thought about the assessments implemented within the "new" curriculum. Whilst most (65%) trainers felt that WBAs were useful to higher grade trainees this dropped to only 30% feeling it was useful for more junior trainees. Another important finding was the agreement that assessments should be varied in number according to each individual trainee, their needs and their level of training.

The free text comments were extremely revealing from this study. They informed that a lack of time was the overwhelming obstacle to their usage, but also a concordance with findings from other studies that inappropriate completion rendered any assessment redundant and that WBAs added to an ever-growing volume of "paperwork" for consultants.

An increasing importance is placed on trainers being properly recognised[88]. The findings from this study provides evidence towards ensuring that in the future trainers should have appropriate time built into job plans such that they can maintain a high standard of training. Surgical training can never move completely away from an apprenticeship model- the very nature of being in theatre with a consultant and performing an operation ensures this will continue to be the case, but there needs to be provision for time to ensure non-operative skills are taught and assessed.

4.4 Case Based Discussions: UK Surgical Trainee Perceptions. Paper and Commentary

Phillips A, Lim J, Madhavan A, Macafee D. Case-based discussions: UK surgical trainee perceptions. *Clin Teach*. 2016 13(3): 207-129[7]

Case-based learning

Case-based discussions: UK surgical trainee perceptions

Alexander Phillips[1], Jeffrey Lim[2], Anantha Madhavan[3] and David Macafee[3]

[1]Department of General Surgery, Royal Victoria Infirmary, Newcastle upon Tyne, UK
[2]Department of General Surgery, Gloucestershire Royal Hospital, Gloucester, UK
[3]Department of General Surgery, James Cook University Hospital, Middlesbrough, UK

SUMMARY

Background: An increasing emphasis on accountability led to the development of the Intercollegiate Surgical Curriculum Project (ISCP) in the UK. A major feature of ISCP was a focus on competence with the institution of formative assessments to aid learning and provide portfolio evidence. Case-based discussions (CBDs) are one of the main formative assessments used at all stages of training. The aim of this study was to review the use of CBDs by surgical trainees to determine if and when they are useful, and whether they are perceived as being used correctly.

Methods: Semi-structured interviews were carried out with both higher and core surgical trainees. Inductive reasoning principles were used to analyse and interpret the responses to open questions. Common themes were determined and thematic analysis was carried out.

Results: Forty-two surgical trainees (21 core and 21 higher trainees) were interviewed. Core trainees felt that CBDs were more likely to be used correctly, and both groups thought that they were a positive feature of training. Few stated that they were used to shape training needs. Positive themes identified included the provision of feedback, identifying learning portfolio evidence and encouraging reflection. Negative themes included a 'tick-box' mentality and that the value was diminished by a lack of engagement with the process from trainers.

Conclusion: Trainees regarded CBDs as a positive feature allowing the discussion of complicated cases, and encouraging higher thinking and reflection; however, concerns were raised regarding their implementation, which has led to a diminishing of their value.

Case-based discussions are one of the main formative assessments used at all stages of training

There is little information regarding CBD validity for surgical training

INTRODUCTION

Since 2007, case-based discussions (CBDs) have been one of the key workplace-based assessments (WBAs) within the Intercollegiate Surgical Curriculum Project (ISCP).[1] The need for WBAs was multifactorial: the implementation of the European Working Time Directive in the UK led to a reduction in learning opportunities, and a more robust method of assessment was required because of increased public scrutiny and emphasis on patient safety.[2,3]

The basic premise of the CBD is an in-depth discussion between a trainee and a trainer to explore the trainee's clinical knowledge and reasoning around a specific topic or patient encounter.[4] As assessment tools, CBDs explore analysis, synthesis and evaluation (the higher elements of Bloom's taxonomy) by encouraging trainees to reflect on learning and to identify development needs. CBDs are reported to have 'face validity' by 'direct observations of workplace tasks', and 'content validity' by being 'blueprinted against all the standards of *Good Medical Practice*', which are the General Medical Council's standards of the duties of a doctor.[5,6]

Although validity research on CBDs for doctor specialities has been performed, there is little information regarding CBD validity for surgical training.[7] With this lack of evidence, the aim of this study was to evaluate the use in practice of CBDs by trainees, and their perceived educational acceptability and utility.

METHODS

Data collection

Recruitment was from surgical trainees in two deaneries (regional geographical organisations for postgraduate training) in the UK. Higher general surgery trainees from all subspecialties were included. A total of 25 higher surgical trainees (HSTs) were contacted, and 21 chose to participate in the study. Core surgical trainees (CSTs) came from a wider range of surgical specialties, including paediatric surgery, otorhinolaryngology, maxillofacial surgery, plastic surgery and general surgery. Twenty-one core trainees were contacted, and all chose to participate. Participants were recruited equally from each deanery.

Interviews were conducted by the same member of the research team. A semi-structured interview guide was used to gather views on: the usefulness of CBDs; the positive features of CBDs; how they are used to shape learning needs; and good and bad experiences of CBDs. The interviews averaged 17 minutes, were audiotaped, and transcripts were verified by the interviewer and another member of the research team.

Data analysis

Interview data were analysed using thematic analysis.[8] Initial coding was carried out by two researchers (AWP and JWL) by identifying the prevalent themes inductively.[9] A subsequent review of themes was carried out using a constant comparative method of analysis to refine emerging themes. Once a comprehensive code structure had been determined, each transcript was re-coded.

RESULTS

A total of 42 surgical trainees were interviewed, with equal numbers of CSTs and HSTs. Interviewee responses demonstrated that a greater number of CSTs than HSTs felt that CBDs were used correctly, and an equal number (29%) felt that they were used to shape training needs. Nine (43%) CSTs compared with 13 (52%) HSTs felt that the usefulness of CBDs depended on the grade of the trainee, being more useful at junior levels of training.

Three broad themes emerged during the analysis that reflected the important role that CBDs can play. These included: teaching and feedback; mentoring; and development. Sample comments from each of these areas by both CSTs and HSTs are shown in Tables 1 and 2.

Teaching and feedback

The teaching theme was frequently coded, and demonstrated that trainees valued CBDs to allow them to question their trainer, and to be questioned and to explore themes. This was coupled with the idea that they could be used to revise topics as well as to provide a base for learning new material. Indeed, trainees appeared to relish the opportunity that CBDs gave for discussing a case in greater depth than they would normally have the opportunity for; however, there was some concern from respondents that discussions could be rushed, which detracted from this opportunity.

Feedback was another theme to emerge. Although not always positive, it was provided with a caveat that it should be honest and constructive; however, there was also a concern from some that the feedback received was rarely useful (Table 1). Core trainees were generally more positive about the feedback provided, although there was an indication that the 'style' of individual trainers impacted heavily on the feedback provided.

Mentoring

Trainees felt that CBDs gave an opportunity to sit with mentors and have a one-to-one discussion about a case. There was real value placed on the 'protected' opportunity to have a senior colleague advise them and act as a 'mentor'. This was particularly

Table 1. Comments from core surgical trainees in each domain

Theme	Dimension	Comment
Teaching and learning	Questioning	'[CBDs] allow one to think beyond the textbook'
	Questioning	'…allow you to discuss a case in depth and look at all aspects of patient care'
	Questioning	'A CBD on diverticulitis with an atypical clinical presentation. The atypical presentation encouraged logical "outside the box" thinking'
	Feedback	'CBDs allow constructive feedback. It is best from a consultant who appreciates your grade. Registrar feedback tends to always be positive, but consultants will tell you how it is.'
	Feedback	'CBDs rarely provide useful feedback, they are done as a task rather than to promote learning, due to lack of time'
	Feedback	'Colorectal clinic – consultant sat in for history and exam and feedback. Excellent quality of feedback, went over all aspects of exam, history and differential diagnosis. Seemed to know what was expected and what appropriate feedback was.'
	Feedback	'It allowed protected feedback time to discuss the boundaries of knowledge base.'
	Learning	'Allows you to discuss a case into depth and review all aspects of patient care and going back to basics of physiology, anatomy, pathology and ethics.'
	Learning	'Consultants often direct me to certain papers or areas that I need to read up on following a case-based discussion'
	Learning	'Through CBD I was able to not only discuss basic medical management but revise ethical and safeguarding issues.'
Mentoring	Time	'I think they are also an excellent opportunity for juniors to actually sit down with their trainers and gauge what level they are at and where they need to be heading.'
	Time	'The eventual comments logged on ISCP are irrelevant, the value is in the time spent discussing the case.'
	Planning	'CBDs provide the opportunity to…identify future learning needs'
	Assessor	'…some consultants are not interested/do not seem to know ISCP very well'
	Assessor	'the Consultant took history and formulated management plan and told me that counted as a CBD'
Development	Evidence	'CBDs provide some attempt at facilitating learning and providing required evidence'
	Evidence	'When minimum number to reach before appraisal CBDs can become a "tick box" exercise with no attached educational value at all.'
	Evidence	'…they are often done out of necessity rather than as an educational tool, therefore they are ruined'
	Reflection	'They [CBDs] prompt the opportunity to reflect on interesting/difficult cases'
	Reflection	'…informed how well you are performing to what is expected'
	Reflection	'In acute settings they have been useful as they provide a way to reflect and reinforce on clinical situations and what was learnt.'
	Reflection	'Discussion of a case that I thought I had managed well, turned out that I could have done better. I learnt from the experience.'

There were mixed views by trainees regarding the impact of CBDs on their development

true with CSTs, whereas HSTs often felt that CBDs were a formalisation of discussions that would naturally occur.

The CBDs also gave trainers a good insight into the trainee's ability, and hence could be used to plan learning needs appropriately. Although CBDs are formally required in the ISCP to demonstrate competencies and improvement, trainees perceived that the way in which CBDs were conducted varied, with some trainers using it as a 'mini-viva'.

Development

There were mixed views by trainees regarding the impact of CBDs in their development. Frequently CBDs were regarded as a 'tick-box' exercise used to fulfil requirements stipulated by training committees. This was further

The CBDs encouraged reflection, which was valued by all training grades

Table 2. Higher surgical trainee comments

Theme	Dimension	Comment
Teaching and learning	Questioning	'CBDs allow for the development of closely associated and integrated clinical and theoretical knowledge'
	Questioning	'…they are usually done in a rushed or ad-hoc manner'
	Learning	'I have undertaken CBDs on emergency cases and then identified that I should do a trauma course in coming years. I have also identified areas of weakness in communication through a CBD which I developed by doing a communication course recommended by a trainer'
	Learning	'A CBD when I looked at a case of necrotic pancreatitis with a consultant in upper GI. We went through the case and then talked about the field generally. My knowledge was examined but I did not feel as if I was in a viva. This left me at ease and more willing to take on new knowledge.'
Mentoring	Time	'As discussions are usually initiated by myself it doesn't add anything to my education as they simply reflect discussions that would naturally occur'
	Assessor	'They are very variable in their usefulness for training and education…dependent on the teaching style of the trainer. CBDs can range from mini-tutorials which are interactive to lectures which are Socratic and idiosyncratic at the same time'
	Assessor	'…a number of trainers used CBDs only on cases where they had criticism…I did learn…but did not find CBDs useful when the trainer adopts a purely viva style assessment'
	Assessor	'The most complicated cases are discussed anyway. This is just a way of formalising it so any future blame goes to the assessor'
Development	Evidence	'Done as a tick box exercise rather than to actually work though a case with a boss'
	Evidence	'They are carried out as a tick box exercise where are rarely completed correctly'
	Evidence	'A bad experience was when I used cases "discussed" in clinic as CBDs. These cases where straight forward and less challenging for my level and it did not involve actual discussion to enrich knowledge.'
	Evidence	'They encourage you to reflect on a case with a senior colleague and benefit from their experience and knowledge in this area.'
	Reflection	'They allow reflection and may influence a change of practice'

supported by concerns that they were sometimes completed incorrectly, which negated their value, and that this resulted from both trainers and trainees not fully engaging with the process. The stipulation of minimum requirements each year encouraged this problem and contributed towards incorrect use, with trainees carrying out CBDs in an unbalanced fashion, mostly towards the end of a placement, in order to fulfil these criteria.

Both groups of trainees commented that CBDs encouraged reflection of their own and other people's practice, which subsequently lead to improvement. The actual process of discussing a case meant that some trainees could identify areas of improvement themselves as well as having them pointed out by the trainer.

DISCUSSION

Case-based discussions (CBDs) are used throughout surgical training and provide evidence for trainee progression. CSTs and HSTs had many similar views, agreeing that CBDs were rarely used correctly and that the emphasis was on the trainee to initiate the assessment. Most trainees felt that CBDs were a positive feature of training: this was more pronounced at the CST level. Both cohorts stated that CBDs were rarely used to shape training.

The HSTs felt that CBDs potentially had more benefit during the junior stages of training, and that this may be because HSTs will frequently have more one-to-one contact with their consultant through theatre and clinic sessions than the CSTs, who instead rely on more senior trainees for their learning. This would explain the extra value that CSTs place on these assessments.

The CBDs encouraged reflection, which was valued by all training grades, and help to provide an insight into learning

needs and personal development. Despite these positive themes, however, a recurring theme was that CBDs were frequently viewed as a 'tick-box' exercise.

Improvements in use

Whereas CBDs are viewed positively, there is a consensus that a good CBD is trainer dependent. Trainers must engage with the process or their benefit is lost. The General Medical Council suggested that WBAs should evolve into 'supervised learning events', and as such trainees need to 'recognise the importance of feedback when learning new skills and that learning should be learner-led'.[10] Although this is certainly a pertinent point, it is also imperative that trainers embrace the importance of constructive feedback and are enthusiastic about using any formative tool. Furthermore, it is important that trainers encourage reflection.[11] One method might be for trainers to initiate a minimum number of WBAs during each post. Alternatively, modification of the current tool into a 'mini-CBD', which would be used to record everyday discussions about patient diagnoses or management, may increase use. These could be recorded by the trainee, thereby reducing the time imposition on trainers, and could serve to provide a broad spectrum of evidence for annual reviews.

There also remains the question of the appropriateness of this assessment. CBDs are useful for discussing complicated cases in depth at high levels of training, and can also be carried out on a more basic level for junior trainees. This is reflected in the responses achieved that were coded in 'Teaching and feedback'. Placing the emphasis on the trainers may reduce the incidence of CBDs being carried out to fulfil quotas and enhance their overall value. As the trainee comments reveal – not all CBDs are equal – so perhaps the quality of CBDs should be evaluated at a trainee's annual review, and trainer-initiated CBDs will help to address this. Further advantages of placing the onus on the trainer include the appropriateness of case selection, and challenging trainees who would no longer be able to carefully select cases that they feel they will be able to do well in.

Imposing a minimum number of CBDs per year, coupled with their generic format, led to many trainees regarding CBDs as a 'tick-box' exercise. Rather than reviewing the number of assessments at annual reviews, it may be beneficial for all WBAs to be reviewed by the trainee's educational supervisor to help inform them about a trainee's progress. This may make the assessments more relevant as they will then be used by a supervisor responsible for, and familiar with, that trainee, rather than by a remote panel.

There are a number of limitations of this study. Trainer views were not solicited, which would complete the picture. Furthermore, although trainees were asked about whether CBDs were used correctly, exploration into whether the trainees actually understood how they should be used would have been helpful: it would have been interesting to see if the attitudes of trainees correlated with their actual performance.

CONCLUSIONS

Surgical trainees believe that CBDs are potentially useful and can aid learning. Perceived misuse, and a lack of engagement by trainers, means that they are often regarded as a 'tick-box' exercise carried out to fulfil a requisite number. Currently their greatest value appears to be in encouraging reflection, exploring clinical problems in greater depth and providing valuable one-to-one time with trainers.

CBDs are useful for discussing complicated cases in depth at high levels of training

REFERENCES

1. Phillips AW, Madhavan A. A critical evaluation of the Intercollegiate Surgical Curriculum and comparison with its predecessor the 'Calman' curriculum. *J Surg Educ* 2013;**70**:557–562.
2. Eardley I, Bussey M, Woodthorpe A, Munsch C, Beard J. Workplace-based assessment in surgical training: experiences from the Intercollegiate Surgical Curriculum Programme. *ANZ J Surg* 2013;**83**:448–453.
3. Carr S. The Foundation Programme assessment tools: an opportunity to enhance feedback to trainees? *Postgrad Med J* 2006;**82**:576–579.
4. Intercollegiate Surgical Curriculum Programme. *Case Based Discussion*. 2010. Available at http://www.iscp.ac.uk/surgical/assessment_cbd.aspx. Accessed on 1 December 2013.
5. Intercollegiate Surgical Curriculum Programme. *Overview of the Assessment System*. 2012. Available at http://www.iscp.ac.uk/surgical/assessment_overview.aspx. Accessed on 1 December 2013.
6. General Medical Council. *Duties of a doctor*. 2015. Available at http://www.gmc-uk.org/guidance/good_medical_practice/duties_of_a_doctor.asp. Accessed on 25 February 2015.
7. Johnson G, Booth J, Crossley J, Wade W. Assessing trainees in the workplace: results of a pilot study. *Clin Med* 2011;**11**:48–53.
8. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;**3**:77–101.
9. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. *Health Serv Res* 2007;**42**:1758–1772.
10. General Medical Council. *Learning and assessment in the clinical environment: the way forward*. 2011. Available at http://www.gmc-uk.org/Learning_and_assessment_in_the_clinical_environment.pdf_45877621.pdf. Accessed on 1 January 2014.
11. Finucane PM, Barron SR, Davies HA, Hadfield-Jones RS, Kaigas TM. Towards an acceptance of performance assessment. *Med Educ* 2002;**36**:959–964.

Corresponding author's contact details: Mr Alexander Phillips, Department of General Surgery, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne, NE1 4LP, UK. E-mail: awphillips@doctors.net.uk

Funding: None.

Conflict of interest: None.

Acknowledgements: None.

Ethical approval: Ethical approval was sought from the Newcastle Hospitals Ethics Committee. It was deemed unnecessary as the study was regarded as service evaluation.

doi: 10.1111/tct.12411

This study reviewed the use of CBDs within surgical training. Whilst previous investigations into formative assessments of the ISCP had been questionnaires of large numbers of trainees providing answers to closed questions, this involved a series of semi-structured interviews with higher and core surgical trainees. Such interviews, with 42 trainees, produced a vast volume of data for thematic analysis.

4.4.1 How could WBAs be improved?

The results from this study suggest that broadly CBDs are viewed as a positive aspect for training and certainly are regarded as useful for surgical trainees. There appears to be a spread of views on whether or not their usefulness is dependent on stage of training, but there is most definitely a consensus that a good CBD is often dependent on the assessor.

There are a number of possible methods that could be implemented to try to improve and enhance their potential benefits. One recurring theme, that was suggested by trainees, was to place a greater emphasis on carrying out CBDs on the assessors. Currently there is a minimum requisite number of CBDs that need to be carried out each year by trainees and if a similar minimum number was placed on trainers this might encourage and foster more acceptance of the tool.

There also remains the appropriateness of this assessment. CBDs are a useful tool for discussing more complicated cases in depth at high levels of training, and also can be carried out on a more basic level for more junior trainees. Placing more emphasis on the assessors may reduce the incidence of CBDs being carried out for the sake of fulfilling an arbitrary number and enhance their overall value.

The imposition of a minimum number of CBDs per year coupled with the generic format of the CBD forms has led to many trainees regarding CBDs as a tick box exercise- often completed in a rush to satisfy the requirements for an upcoming ARCP. Rather than purely reviewing numbers at the ARCP it may be beneficial if all WBAs were reviewed by a trainee's educational supervisor and used to help inform them about a trainee's progress. This may make the assessments more relevant as they are being used by a supervisor responsible for that trainee, and who will be largely aware of their strengths and weaknesses, rather than by a remote panel, whose main aim is to review quantity of WBAs completed rather than quality.

Arguably the single factor that will improve CBDs and indeed all WBAs is a greater belief by trainees and assessor that CBDs are beneficial to training. This will almost certainly go hand-in-hand with their prudent and appropriate use.

4.4.2 Strengths and Limitations

4.4.2.1 Numbers and Geography

This study evaluated the views on CBDs by Core and Higher Surgical Trainees within two deaneries in the UK. It could be argued that those involved in the study represent only a small proportion of both sets of trainees nationally. It is also possible that the employment of WBAs may differ between deaneries (although it could be argued it will differ between hospitals and individuals). Conducting telephone interviews, with such a large number of doctors, required considerable time. This was a major undertaking in recruitment, actually carrying out interviews, and analysis of a large volume of data. Since those questioned at CST level came from an array of subspecialities and had worked in different hospitals, it would seem reasonable to extrapolate their view points as being largely representative of those shared with other CSTs. The HSTs involved had worked at multiple sites within their deaneries and their experiences were based on having worked at a number of different hospitals. All had carried out many CBDs during their training and thus their answers will not have been based necessarily on what has happened at their current place of work. All the HSTs were General Surgical Trainees, but no consideration of which subspecialty they were currently working in or where their future intentions lay was recorded.

4.4.2.2 Methodology

Although there are weaknesses in using a telephone questionnaire for data collection, this method allowed the most views and attitudes to be surveyed and therefore was considered to be the best method for gaining answers to the original questions. The other potential methods that could have been employed were focus groups, face-to-face interviews and postal or email questionnaires. Focus groups, may have yielded good results, but they demand the physical and simultaneous availability of participants and logistically it may have been more difficult to recruit to these. Further, althought they allow the topics to be discussed at length, some individuals may be concerned about voicing

opinions in front of peers. Carrying out individual questionnaires allowed independent views to be gauged and possibly for people to speak more candidly about their experiences. Face-to-face interviews would arguably have been the best method of collecting data, allowing for depth of discussion and also providing an insight into responses by including body language and being more personal. However, this again would have been logistically very difficult to achieve given the work commitments of those recruited and the wide geographical area the participants were in. Telephone questionnaires thus seemed a reasonable compromise, as it negated geography as an issue and allowed a convenient time for interviews to be conducted more easily. The final option considered, email or postal questionnaires, may have allowed a greater number of participants to be surveyed, but this may have compromised the depth and level of responses.

4.4.2.3 Questionnaire

It is important to address the issues of reliability and validity of the questionnaire. Reliability was not formally assessed with this questionnaire. According to Stenbacka, "*...the concept of reliability is even misleading in qualitative research. If a qualitative study is discussed with reliability as a criterion, the consequence is rather that the study is no good*"[89]. To counter this Lincoln & Gubba suggest that the term "dependability" should be used in qualitative research to replace reliability[90]. In this instance no attempt at test- retest reliability was carried out although the questionnaire was piloted amongst external trainees to ensure that appropriate responses were received. However, the questionnaire design comprised a set of questions with discrete answers, followed by pre-conceived questions which were consistent for each trainee. Oppenheim states that the wording of a question is important when asking about attitudes[91]. Further, Silverman placed importance on each interviewee understanding the question in the same way[92]. As only one person (myself) conducted the interviews this should have led to consistency in how questions were asked and how they were understood; if there was any confusion to the actual meaning of the question the intent could be easily clarified. Returning to Lincoln and Gubba's notion of dependability, having a single interviewer, familiar with the intent of the questions, and asking each interviewee in the same fashion, should lead to an increased dependability of results obtained. One suggestion for achieving greater validity is to minimise the amount of bias as much as possible[93]. On reviewing the questionnaire, it would seem reasonable to conclude that the

questions asked are designed to answer the initial questions. The problem that remains is thus whether bias will occur because of the way that interviewees answer the questions or because of a misinterpretation by the interviewer and latterly by the person responsible for analysing the data.

It may have been possible to strengthen the questionnaire when carrying out the pilot study by using the questions, and then revisiting those interviewed with the same questions to determine if a consistency existed. However, the open nature of the questions in the second half of the questionnaire led to the emergence of themes rather than a vast disparity of data which could be used to make the commonsense argument that the questions involved do aid in finding answers to the original aims.

4.4.2.4 Bias

There are a number of potential areas for bias within this study which could consequently influence the validity and reliability of the results obtained. The bias within this study may be related to the interviewer, handling of the data and expectations of the interviewee. It is possible for the interviewer to influence results with their own attitudes and opinions. This, at its worst, may lead to seeking responses that support preconceived ideas. The initial aim was to be as neutral as possible and it was unclear from the outset what the attitudes of my peers were likely to be. It is possible that bias could also be introduced by misunderstandings- by the interviewer of what the respondent is saying, and by the respondent of the actual underlying meaning of the question. Further, respondents may also feel that a specific answer is "expected" of them. Whilst all interviews were anonymised and assurances of confidentiality were given, there is still a chance that respondents may not be completely honest.

4.4.3 Grounded Theory versus thematic analysis for qualitative methodology

The methodology in this study involved a thematic analysis rather than a grounded theory approach. The two approaches differ, and employing grounded theory is a challenging approach to determining outcomes from qualitative data. Grounded theory is a "*qualitative research method that uses a systematic set of procedures to develop an inductively derived theory about a phenomenon*"[94]. Once data has been collected an iterative process is carried out which involves coding data, comparisons between data collected often involving diagrams and memos and re-analysing the original data using the emerging themes. As an iterative procedure, the process is carried out repeatedly until data "saturation"

occurs. Eventually a model is created to link relationships between the codes. This "theory" usually involves a set of concepts which are related to each other and explain the phenomenon occurring.

Thematic analysis has some similarities, in that it involves coding the data and looking for themes but involves a simpler methodology which aims to find themes within the qualitative data rather than an overarching theory. It involves familiarisation with the data and then the production of initial codes. Once all the data has been coded a search for themes amongst the codes is carried out. Some codes may form main themes, or be sub-themes. Following this the original data is revisited to consider it in relation to the themes produced. These themes are then defined and named providing a final thematic grid.

4.5 Entrustable Professional Activities and Competencies in Practice: A new paradigm

The idea of "Entrustable Professional Activities" was originally suggested in 2005. It has been developed as a method of identifying particular activities, or components of procedures that trainees may demonstrate they are able to perform independently[95]. Within the UK this concept has been entitled a "competency in practice" (CiP). The idea is that the unit of activity or CiP can be devolved to a trainee once competence has been achieved[96].

EPAs were developed to bridge the gap between competency based training and clinical activities. EPAs should thus be regarded as units of clinical work or task. The decision to allow a trainee to perform a specific task is called an entrustment decision[97]. These events are part of everyday activities and usually occur in an ad hoc fashion. These occur when the trainer feels that the trainee is capable of the managing the complexity of the task and the risk in allowing the trainee to perform the task are acceptable. EPAs involve linking of multiple competencies simultaneously and can be used to assess whether a trainee is able to perform a "real-world" task. In essence EPAs are units of professional practice whilst competencies describe specific abilities (knowledge, or technical ability, or communication ability)[98].

It is likely that EPAs/ CiPs will be increasingly used to provide evidence that trainees have met the required standards for autonomous practise. They will be used in addition to the

WBAs already in place, and serve to provide a method for demonstrating that achievement of competencies can be linked to provide real world patient care.

4.6 Student Views on Two Styles of Feedback. Paper and Commentary

Nesbitt CI, Phillips AW, Searle RF, Stansby G. Student views on the use of two styles of video-enhanced feedback compared to standard lecture feedback during clinical skills training. *J Surg Educ.* 2015 Sep-Oct;72(5):969-73[8]

Student Views on the Use of 2 Styles of Video-Enhanced Feedback Compared to Standard Lecture Feedback During Clinical Skills Training

Craig Nesbitt, MRCS, MD, Alex W. Phillips, FRCS, Roger Searle, PhD, and Gerard Stansby, MA (Catab), MB, MChir, FRCS

Colorectal Surgery, Sunderland Royal Hospital, Sunderland, United Kingdom

BACKGROUND: Feedback plays an important role in the learning process. However, often this may be delivered in an unstructured fashion that can detract from its potential benefit. Further, students may have different preferences in how feedback should be delivered, which may be influenced by which method they feel will lead to the most effective learning. The aim of this study was to evaluate student views on 3 different modes of feedback particularly in relation to the benefit each conferred.

METHODS: Undergraduate medical students participating in a surgical suturing study were asked to give feedback using a semi-structured questionnaire. Discrete questions using a Likert scale and open responses were solicited. Students received either standard lecture feedback (SLF), individualized video feedback (IVF), or enhanced unsupervised video feedback (UVF).

RESULTS: Students had a strong preference for IVF over UVF or SLF. These responses correlated with their perception of how much each type of feedback improved their performance. However, there was no statistical difference in suturing skill improvement between IVF and UVF, which were both significantly better than SLF.

CONCLUSION: Students have a strong preference for IVF. This relates to a perception that this will lead to the greatest level of skill improvement. However, an equal effect in improvement can be achieved by using less resource-demanding UVF. (J Surg Ed 72:969-973. © 2015 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: video-enhanced feedback, clinical skills training, students views, suturing

COMPETENCY: Medical Knowledge, Practice-Based Learning and Improvement, Patient Care

Correspondence: Inquiries to Craig Nesbitt, MRCS, MD, Healeyhope Barn, Waskerley, Consett, Co Durham, DH8 9DB; e-mail: craigiainnesbitt@gmail.com

INTRODUCTION

Feedback plays a vital role in enhancing learning and providing the optimum opportunity for students to benefit from a teaching environment. Feedback is often delivered in an ad hoc fashion and can subsequently be ineffective.[1] In the teaching of practical skills, students who have difficulty may be provided with individual feedback, but where large groups are involved, most of the students may obtain only generic or unstructured feedback.[2]

Previous studies have suggested that feedback immediately after performing a task is most beneficial, and how the feedback is provided and its content are also key components.[3] Further, it has been established that students value feedback and are aware of the effect that it may have on their learning.[4-6] The recent national student survey[7] revealed that across the UK, and in all undergraduate curricula, students are unhappy with the amount of feedback they receive from their respective faculty, yet most demonstrate good insight and empathize at the difficulties teachers encounter at providing effective feedback.[2]

Both teachers and students recognize that time and resources are limiting factors, which can make individualizing feedback difficult. However, students have stipulated that this difficulty could be alleviated by more engagement by their tutors[8] and more consistency.[9] Rowe and Wood[2] postulated that large class sizes may lead to some of the constraints experienced.

Understanding students' preferences for feedback can help obtain acceptance and optimize the learning experience. However, there have been few studies investigating student attitudes to different feedback methods, particularly in relation to the effect that the feedback subsequently has on attaining a skill.

The aim of this study was to review medical undergraduate students' opinions on 3 different feedback methods after being initially taught how to execute a simple suturing skill.

  1931-7204/$30.00 *http://dx.doi.org/10.1016/j.jsurg.2015.04.017*

METHODS

Undergraduate medical students from Newcastle University were invited to participate in a randomized clinical trial comparing 3 types of feedback on learning a basic suturing task. A power calculation (recently published[10]) determined a minimum requirement of 30 subjects (10 in each arm).

Participating novice students were taught to perform an instrument-tied reef knot by watching a standardized video. After performing the task, students were randomized to each of the 3 feedback groups using a closed envelope system. Group 1 received standard lecture feedback (SLF), Group 2 unsupervised video feedback (UVF), and Group 3 individualized video feedback (IVF). They were then invited to repeat the task before filling in the feedback questionnaire. This involved rating 3 statements on a Likert scale.

1. "The feedback I received was adequate"
2. "The feedback I received improved my subsequent performance"
3. "I would be highly satisfied with this form of feedback for future clinical skills training. For example—following venopuncture, basic life support training, etc."

There was also an opportunity for free text responses regarding their perceived advantages and disadvantages of each form of feedback.

Students suturing performances were scored by experts in real time and also subsequently had their recordings scored by 2 further experts who were blinded to the candidate and whether the performance was pre- or postfeedback. Experts used a validated Objective Structured Assessment of Technical Skill[11] scoring tool, which was adapted to this basic suturing task.

Feedback Styles

Students were randomized to 1 of 3 feedback methods before repeating the suturing exercise.

Group 1: SLF: 20 minutes

Students were given a generic didactic lecture lasting approximately 20 minutes. This covered most common errors and pitfalls and students were permitted to ask questions although care was taken not to "individualize" feedback.

Group 2: UVF: 20 minutes

Those receiving UVF reviewed their performance on a laptop in isolation. As well as their own performance they had access to a video of an expert performing the task and a further video of an expert giving hints and tips. Students were allowed to rewind and replay sections of the video as they wished.

Group 3: IVF: 20 minutes

The third cohort reviewed their video performance accompanied by an expert who gave technical feedback specific to their performance. Again candidates and experts were permitted to pause, rewind, and replay the video at any point and to ask technical questions. A maximum of 20 minutes was allowed for this feedback.

RESULTS

A total of 32 students participated in the study.

Students rated their agreement with 3 statements on the posttrial questionnaire on a 5-point Likert scale, with 1 representing the greatest disagreement with the statement and 5 the greatest agreement. The mean scores for each group are show graphically in Fig. 1. These scores are compared statistically in Fig. 2.

Candidates Performance Scores

Candidates' mean overall performance score improved following SLF (39.3-53.7), UVF (38.4-57.8), and IVF (38.6-59.7). This improvement was statistically significant in all cohorts. However, both UVF and IVF demonstrated a statistically significant greater improvement than candidates randomized to SLF. There was no significant difference in improvement demonstrated between candidates receiving UVF and IVF.[10]

Candidates Posttrial Comments

Students were invited to make comments on their perceived advantages and disadvantages of the feedback they received. A selection of these comments is displayed in Tables 1 to 3.

DISCUSSION

The analysis of students postquestionnaire Likert scores indicated that statistically there is a significant difference in favor of video feedback (UVF and IVF) vs SLF when

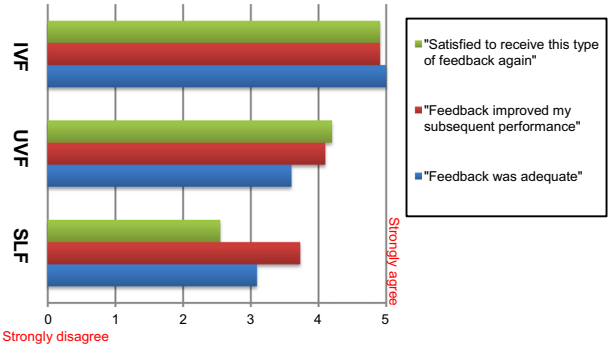

FIGURE 1. Mean score for agreement with statements regarding trial feedback.

Feedback / Group / p-value	*"The feedback I received was adequate"*	*"The feedback improved my subsequent performance"*	*"I would be satisfied to receive this type of feedback again"*
1 vs 2	0.009	0.274	0.270
1 vs 3	0.000	0.001	0.000
2 vs 3	0.202	0.020	0.002

[1]Mann Whitney U Test (Monte Carlo Sig 2-Tailed Test)

Statistically significant

FIGURE 2. Significance test of between-group difference, comparing posttrial questionnaire Likert scores.

considering the statement that their feedback was "adequate." However, further scrutiny shows that students prefer one-to-one, face-to-face feedback, and the IVF group rated all statements significantly higher compared with SLF, and higher than UVF in terms of "improving their subsequent performance" and "satisfaction to receive this type of feedback again" (Figs. 1 and 2).

This observation is also reflected in students' free text responses. Students perceive that individualized feedback yields the greatest benefit, claiming that it is "*tailored to specific weaknesses*" offering a "*chance to ask questions, ability to watch and see mistakes and then be told ways to improve.*" Intuitively a candidate acknowledged the feasibility of it during large group teaching sessions: "*I don't think the medical school has the time/money to offer this sort of feedback.*"

Despite students' apparent preference for individualized feedback, they were also satisfied with UVF, and appreciated the opportunity to "*compare what you do to the 'pro' and you can pick up on the mistakes you make, very useful.*" Students were dissatisfied with SLF, they felt it "*didn't highlight specifically what I was doing wrong & I feel that I could continue making the same mistakes" and "lecture format was dull meaning I occasionally stopped paying attention"* (Table 1).

Despite students' apparent belief that individualized feedback is superior, the clinical performance scores do not support this assertion; IVF did not yield any significant improvement vs UVF.[10] This observation is supported by the conclusions of O'Connor et al.,[12] who showed no significant improvement in laparoscopic suturing performance when candidates received additional expert feedback compared to simply the knowledge of their performance score.

The teaching of surgical skills has traditionally followed an apprentice model with trainees frequently receiving one-to-one feedback from their trainer particularly when performing operations. However, pressures on training opportunities and the European Working Time Directive have meant an increased focus on the use of simulators to learn skills and a greater ratio of trainees to trainers.[13] Being able to scrutinize one's own performance with assistance of expert videos may help enhance trainee opportunities for attaining skills and provide an alternative to the more traditional learning model.

Retrospective self-assessment, the ability to rate one's performance on a recent exercise,[14] is vital if learning is to

TABLE 1. Candidates' Perceived Advantages and Disadvantages of SLF

Advantages	Disadvantages
It did highlight some common mistakes	It was not as helpful as personalized feedback
Highlighted important areas for improvement	Did not highlight specifically what I was doing wrong and I feel that I could continue making the same mistakes
Reminded me of errors I had made, provoking some internal reflection into my own performance	It was too generalized
Ensures that everyone receives some form of feedback	It would have been much better if personalized
Can be given to large groups	Lecture format was dull meaning, I occasionally stopped paying attention

TABLE 2. Candidates' Perceived Advantages and Disadvantages of UVF

Advantages	Disadvantages
Allowed you to see if you had forgotten anything, watching the expert one reminded you of anything you had forgotten	No verbal feedback on performance given or written so you might not pick up on everything still
You get to compare what you do to the "pro" and you can pick up on the mistakes you make, very useful	Watching my own video was of limited use as all I could see was that I clearly had no idea what I was doing
People are naturally self-critical and so giving people the opportunity to observe themselves is both time efficient and beneficial	Did not know if I was doing anything wrong without noticing, direct feedback would help here
I was able to identify the mistakes I made and hence correct these in subsequent attempts	Maybe linking the video with examiners feedback would make the quality of feedback even more helpful to students
It was very useful being able to watch my performance, especially the fact that I could see my mistakes	No personal feedback from the markers

TABLE 3. Candidates' Perceived Advantages and Disadvantages of IVF

Advantages	Disadvantages
The video camera was very useful in showing where I went wrong. Questions posed by the examiner also helped my understanding	Listening to other peoples questions/feedback may be useful in gaining extra information
You can see exactly where and when to improve	Took a bit of time
Because feedback was individual I got a lot of information about my own performance/technique	Time consuming
Feedback was relevant to the individual task, whereas watching the video helped the feedback to be more specific	I do not think the medical school has the time/money to offer this sort of feedback!
Tailored to specific weaknesses, chance to ask questions, ability to watch and see mistakes, and then be told ways to improve	You cannot offer this in a large class, puts you under quite a lot of pressure

occur where individual feedback cannot be provided. There have been suggestions that this can be enhanced by using external benchmarks of performance.[15] Self-review of videos has been shown to enable significant improvement in surgical skills[16] but to ensure that a flawed perception is not developed it should be accompanied by external sources such as peer review, faculty input, and other external modalities.[15]

Feedback methods may be regarded as directive or facilitative—directive feedback tends to be specific in nature, which can result in greater performance improvement, whereas facilitative feedback promotes students in making their own revisions[17] and its nonspecific nature may lead to uncertainty and hinder learning.[18-20] This may explain why students were less satisfied with UVF.

This is a small study reviewing the opinions of students motivated to learning a new skill. It reviews only 3 feedback modalities among the many that exist. It would be beneficial to conduct further studies investigating both undergraduate and postgraduate opinions on the feedback styles investigated and also to determine whether these feedback techniques are held in the same regard when executing other clinical skills.

Based on the results of this trial, students prefer individual tuition when trying to learn new clinical skills. There is an understanding that this is not always practical owing to constraints of time and resources. Students underappreciate the benefit of unsupervised video-enhanced feedback. The ability and technology for recording one's own clinical performance is now readily available in clinical skills departments in the UK[21] and this could be a valuable addition to modern medical learning.

REFERENCES

1. Gran SF, Braend AM, Lindbaek M. Triangulation of written assessments from patients, teachers and students: useful for students and teachers? *Med Teach*. 2010;32(12):e552-e558.
2. Rowe AD, Wood LN. Student perceptions and preferences for feedback. *Asian Soc Sci*. 2009;4(3):P78.
3. Rucker ML, Thomson S. Assessing student learning outcomes: an investivation of the relationship among feedback measures. *Coll Stud J*. 2003;37(3):400-404.
4. Higgins R, Hartley P, Skelton A. The conscientious consumer: reconsidering the role of assessment feedback in student learning. *Stud High Educ*. 2002;27:53-64.
5. Hyland P. Learning from feedback in assessment. Hyland P, Booth A, editors. The Practice of University History Teaching. Manchester: Manchester University Press, 2000. p. 233-247.
6. Weaver MR. Do students value feedback? Student perceptions of tutors' written responses *Assess Eval High Educ*. 2006;31:379-394.
7. National Student Survey: findings and trends 2006-2010 [Internet]. [cited 2012 May 24]. Available from: www.hefce.ac.uk/media/hefce/content/pubs/2011/201111/11_11.pdf.
8. Weston PSJ, Smith CA. The use of mini-CEX in UK foundation training six years following its introduction: lessons still to be learned and the benefit of formal teaching regarding its utility. *Med Teach*. 2014;36(2):155-163.
9. Weller JM, Jolly B, Misur MP, et al. Mini-clinical evaluation exercise in anaesthesia training. *Br J Anaesth*. 2009;102(5):633-641.
10. Nesbitt C, Phillips A, Searle R, Stansby G. Randomised trial to assess the impact of supervised and unsupervised video feedback on teaching practical skills. *Journal of Surgical Education*. pii: S1931-7204(14)00347-X. doi:10.1016/j.jsurg.2014.12.013.
11. Phillips AW, Madhavan A. A critical evaluation of the intercollegiate surgical curriculum and comparison with its predecessor the "Calman" curriculum. *J Surg Educ*. 2008;70(5):557-562.
12. O'Connor A, Schwaitzberg SD, Cao CGL. How much feedback is necessary for learning to suture? *Surg Endosc*. 2008;22(7):1614-1619.

13. Sadideen H, Hamaoui K, Saadeddin M, Kneebone R. Simulators and the simulation environment: getting the balance right in simulation-based surgical education. *Int J Surg*. 2012;10(9):458-462.

14. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *J Am Med Assoc*. 2006; 296(9):1094-1102.

15. Zevin B. Self versus external assessment for technical tasks in surgery: a narrative review. *J Grad Med Educ*. 2012;4(4):417-424.

16. Ward M, MacRae H, Schlachta C, et al. Resident self-assessment of operative performance. *Am J Surg*. 2003;185(6):521-524.

17. Archer JC. State of the science in health professional education: effective feedback. *Med Educ*. 2010;44 (1):101-108.

18. Kluger A, DeNisi A. The effects of feedback intervention on performance: a historical review, a meta-analysis, and preliniary feedback intervention theory. *Pscyhol Bull*. 1996;119:254-284.

19. Goodman J, Wood R, Hendrickx M. Feedback specificity, Learning Opportunities and learning. *J Appl Psychol*. 2004;89(5):809-821.

20. Phye G, Sanders C. Advice and feedback: elements of practice for problem solving. *Contemp Educ Psychol*. 1994;19:286-301.

21. ⟨http://www.scotiauk.com/⟩.

The ability to undertake basic practical procedures safely and effectively is an essential prerequisite for the graduating medical student. In the United Kingdom (UK), the General Medical Council (GMC) sets the standards for undergraduate learning in "*Tomorrow's Doctors*"[99] stating that graduates are expected to "*be able to perform a range of therapeutic procedures*" including intravenous cannulation, urethral catheterisation and skin suturing, amongst some 24 other practical procedures outlined in the guidance[99]. Additionally, the GMC places significant emphasis on the importance of student-directed feedback, where students are accountable for learning and subsequent performance[99].

Indeed, feedback is vital for medical practitioners[100]. Feedback affords learners the opportunity to identify the gap between actual and desired performance levels and subsequently how that gap can be narrowed to improve performance. Medical students in particular have been shown to have increased satisfaction when feedback on their performance is optimised in terms of both quality and quantity[101]. In spite of this, in a national survey of UK-based students across all subject disciplines over nine years, medical students were amongst the least satisfied with feedback provision in terms of punctuality, quality and utility[102]. The continual expectation for educational institutions to strive to improve the quality of feedback, in turn, drives the pursuit of newer feedback modalities, such as the use of video technology in the context of simulated technical skills.

Whilst there are several barriers to learning associated with feedback, emotional distress, or anxiety, may prevent learners from using and accepting the feedback provided to them, especially if the feedback either threatens their self-esteem or is not in line with their own assessment of their ability. Feedback associated with negative emotions may have a long-lasting impact on students' learning[103] and it has been previously acknowledged that excessive levels of stress may have an adverse effect on performance in a simulation and surgical skills context[104–108].

Using recordings to provide feedback is by no means new and being able to self-critique is an important skill. It was important to evaluate how students felt about the three feedback mechanisms used within the pilot study. Given the pilot study had established no significant difference in performance between the two groups able to review their own performance, it is also important to know if those involved found the mechanism of feedback acceptable.

The main questions were kept simple and based on a five-point Likert scale in order to try and tease out whether students were happy with the feedback and whether they thought it had

improved their performance. They were also allowed free-text responses. Students greatly preferred feedback from an expert and valued the one-to-one tuition. However, it was noticeable that several students had insight into the fact that this type of feedback required considerable resources in the form of staff to provide feedback. Self-review of videos was again preferred over the generic feedback. Some of the comments were particularly revealing indicating that without an expert present a level of uncertainty existed that skills were being performed incorrectly.

Ultimately, the best way of utilising these feedback mechanisms may be to involve a blend of them. Further research needs to be done to determine how to best utilise experts when they are available. It may be that expert review can be integrated into the learning pathway after students have a basic appreciation of the skill. Alternatively, it may be that early input of an expert will allow the basics of a skill to be learnt, and students can be told where they are commonly going wrong. They can then use self-review with videos to perfect a skill.

However, what is applicable in simple technical tasks, may not translate to more advanced surgical skills. Further research into both of these needs methods needs to be carried out to try and determine the most effective way of improving learning.

4.7 Chapter Conclusions

The papers in this chapter reveal that many of the concerns that trainees have with WBAs are shared by trainers. Both regarded WBAs as a positive feature of training with the potential to enhance learning, but both recognised there are a number of flaws and a tendency towards not being used correctly. This can translate into them becoming a “tick-box” exercise with limited educational value. Interestingly whilst trainers complain about a lack of time to complete the appropriate assessments, trainees also recognise this is a limitation.

With respect to the video feedback used within the two trials, students preferred to have an expert go through their video rather than to review their own performance with only an expert video as guidance. This is probably due to a lack of confidence, and may change as students become more senior and more accustomed to self-evaluation. It may also be that the greatest benefit from the video technology will arise from the two feedback mechanisms being used in conjunction with each other, and that this will potentially reduce the amount of time an expert

trainer needs to be available, and allow trainees to practise a skill so the steps and performance become second nature.

Both the technology based feedback and WBAs are able to enhance the learning experience by providing beneficial feedback.

Chapter 5 Conclusions

This concluding chapter commences with a reminder of the initial aims of this thesis.

5.1 Aims

1. Analyse the changes that have occurred in surgical training and the reasons behind them
2. Determine that quality training can occur within the current surgical training programme
3. Review the validity and reliability for WBAs within Surgical Training
4. Determine the impact of video-based feedback on learning technical skills
5. Explore the perceptions of users of these feedback mechanisms.

5.1 Discussion

This thesis sought to gain further insight into surgical training within the UK and look at the underlying drivers for change. The massive changes that have occurred have necessitated the implementation of WBAs, and will require the increased use of technology. Technology-based systems that aid with the provision of feedback will no doubt become increasingly important in the future, and the studies within this thesis provide a basis for further research and how it may be applied- both at undergraduate and postgraduate levels.

Importantly, whilst there has been trepidation towards many of the changes in surgical training, both trainers and trainees recognise the potential value of WBAs and trainees are receptive to new methods of receiving feedback.

The introductory chapter of this thesis looked at the changes that had occurred in surgical training over the last 30 years. This led onto the first paper “A Critical Evaluation of the

Intercollegiate Surgical Curriculum and Comparison with its Predecessor the "Calman" Currciulum"[1]. It demonstrated a paradigm shift, with the creation of a curriculum rather than a syllabus, defined learning targets and an aim to link operating with non-technical skills on an online platform. The second paper in this chapter evaluated outcomes from a high volume unit performing oesophagectomy- with a focus on the impact of trainee involvement. Reassuringly this found that trainees performing all or part of such a complex procedure did not compromise patient outcomes. At a time when there have been advances in surgical techniques, and increased focus on patient safety and surgeon outcomes there has been an increasing realisation that quality training is vitally important. Previously volume of cases and time have been the mainstay for establishing that an individual is sufficiently trained, but the idea of "competency" has become an increasing focus. This is not without issues as it could be argued that the focus on achieving competencies detracts from the aspiration to excellence. Tooke in his report *Aspiring to Excellence* introduced the need for a quality assurance process within training[109]. *Time for training,* a subsequent report, highlighted the importance to establish measures to evaluate training outcomes. It determined that effective training could occur in a 48 hour week, but that often training opportunities were lost and training needed to be prioritised in order to minimise this[110].

The next chapter evaluated WBAs within the surgical curriculum. The initial paper highlighted the dissatisfaction and distrust of WBAs[3]. The anecdotal nature of the evidence for their implementation, particularly in an age when evidence based medicine has become integrated into all physicians' minds meant that their true value was rarely appreciated as they were incorrectly used. However, their potential for identifying the struggling trainee is possibly where they will have the greatest impact. The two randomised trials that followed demonstrated the potential impact that new technology may have on learning clinical skills[4,5]. Despite being simple tasks, results were corroborated in two separate studies and the future lies with students being able to use technology readily at their disposal- mobile phones, tablets, to record their own simulated attempts and learn from them. Whilst it could be argued that nothing beats "real-life" for learning, simulators are ever-improving, providing a better approximation of what is required in the real world and can aid students in learning steps and building confidence.

The last chapter explores the perceptions that users have in the preceding feedback methods. There is no doubt that both trainers and trainees realise the importance of feedback. They also appreciate that WBAs can be a useful tool when used appropriately. Students did not realise

quite the impact their own self-assessment had on learning a new skill, and the ability to self-appraise, which is expected at postgraduate level, needs to be introduced as a concept at an early point in the medical curriculum in order for students to maximise their training opportunities.

5.2 Limitations

Many of the limitations with each study have been discussed within the body of each manuscript and there has been further evaluation within each chapter.

The first chapter involved a critical evaluation and comparison between two curricula. It is possible that bias may have been introduced into the findings, as someone who was currently using the new curriculum. However, I was aware of the predecessor and had limited experience with the old curriculum. Drawing comparison between unwieldy concepts such as curricula meant that a systematic approach needed to be applied. This involved looking at each of the different facets of the curriculum and comparing each curriculum's approach to different items such as outcomes, assessment and syllabus.

The second paper, which retrospectively reviewed a prospectively kept database was limited by a disparity in group size. It was also difficult to know the relative experience of each trainee prior to working at the unit. The study did not provide a mechanism for establishing a learning curve, but did establish its primary aim of demonstrating comparable outcomes between each of the groups analysed.

The validity and reliability of WBAs was assessed in the first manuscript of Chapter 3. This was carried out as literature review and systematic review. There was limited data on this topic within surgical training and much of the data needed to be extrapolated from publications in other medical specialities.

The two randomised trials could be criticised for failing to allow students an opportunity to practice a skill. The low sample size in the first study was compensated for by the large numbers in the second study which sought to establish that this form of learning could be transferred to other simple skills. The main question that remains is whether the statistical improvement found translates to an actual improvement in performing the skill? Further research on this is required, as is further work into skill retention, and how effective this type of feedback is for more advanced skills.

Chapter 4 evaluates perceptions of trainers and trainees. The response rate from the surgical trainers study was small- but still provided a snapshot of their views. It is likely that many of the responses will have come from those with an interest in training- introducing self-selecting bias. Despite this a breadth of views was obtained, and the free-text comments helped ascertain many of the problems trainers encountered. This is a topic that requires further investigation and may be better suited to a focus group where problems can be analysed at a deeper level.

An attempt to get a deeper level of understanding of the use of CBDs with semi-structured interviews was used in the next paper[7]. The confines of the style of the journal *Clinical Medicine* meant that only a limited discussion could be included regarding the limitations. Further studies into trainee views on other WBAs may be valuable and it may become apparent that some WBAs are more appropriate at different stages of training than others.

The final paper was limited in the depth of responses that it obtained[8]. Whilst it was valuable to gauge what students thought of the feedback mechanisms employed in the trial, further in-depth discussions either as a focus group, or individual interviews would allow students to discuss the problems they had. It may also help establish improved mechanisms for receiving the feedback. It is hoped that this training tool will be implemented into the undergraduate curriculum to aid learning.

5.3 Future Work and Challenges

As with any research, often it evolves and leads to the formation of further research questions. This is true within all the components of this thesis. Whilst further research may aim to address the limitations that have been highlighted above, the focus needs to be on concepts that will improve surgical training.

One major challenge is in the production of robust evidence. Randomised trials, particularly with students in medical education is often met with concern regarding the ethics of providing one group of students with a potential advantage. This stems from providing different teaching and feedback to each group. The medical school have strong rules regarding the use of students in such studies, and the opportunity to take part must be offered to all. Further, having a control group, which merely repeats a task and receives no feedback and no direct teaching was regarded as unethical during the application for ethical approval.

In the first randomised trial included in this thesis, a standardised lecture was used as a feedback mechanism, and regarded as the "control" group. This was used for comparison to those receiving feedback either from an expert or allowed to review their own performance. In the second study such a "control" group was felt not to be required. This was due to the fact that the initial study demonstrated improvements in all groups but significantly more with those using video technology. Thus the aim was to determine whether any significant difference existed between the video feedback groups.

This ethical issue makes it extremely difficult to develop and expand randomised trials. It should be noted that an intervention cannot necessarily be regarded as beneficial. It is possible that self-review could have strengthened any existing flaws in technique. Whilst it is important to ensure that after any study students get useful feedback (as much to reward the time they sacrifice for participating in such studies), in order for the research itself to be robust it must provide distinct arms for comparison which may potentially disadvantage one cohort at the time of the study.

Certainly further studies into the application of video feedback needs to be considered. The findings from the trials in this study raise the possibility of a resource efficient method for learning skills. Similarly, further work needs to be done with regards to WBAs, their appropriateness at different levels of training, and the benefit of each of the different types of WBAs.

The two randomised trials carried out present a number of unanswered questions in relation to the effectiveness of video feedback in learning skills. Perhaps the most obvious of these questions is which type of feedback leads to the best skill retention. This requires further study. Whilst a similar study that assesses students at fixed time points after "learning" a skill and receiving feedback may help answer this question, it will not provide a real-world answer. In training, students will have the opportunity to practise a skill. Further, the use of videos will allow students to review their performance regularly. Thus a comparison of skill retention with students allowed to review their own performance at will, compared to having a single feedback session, might better represent the options present in actual training.

Another potential questions is how having expert feedback may be used in combination with self-evaluation of videoed performances. Would the best combination of the two involve students receiving expert feedback early on and then using videoed performances to hone a

skill? Or, alternatively, should they try and become comfortable at performing the skill by reviewing their own videoed performances and then have an expert help finesse the performance?

Finally, with regards to video feedback, what is its role in learning more complex skills? This includes surgical procedures, which require several steps- further research needs to be carried out to determine the impact of video feedback and self-review of performances on learning these skills.

WBAs remain controversial tools that have not been fully accepted by both trainers and trainees. The addition of EPAs/ CiPs may be beneficial in providing units of assessment based on actual clinical work. There is however, still much research that needs to be carried out in determining the optimum use for such tools. Their adoption has, arguably, been unstructured in surgical training. Often the overriding aim is to carry out a set number of WBAs, without looking at the quality of the interaction or which tools are most beneficial at each stage of training. Those responsible for the curriculum have made the use of these tools essential and have placed the responsibility for driving their use entirely with trainees. Further research evaluating who should be responsible for driving their use, in order to gain maximum benefit from them, and which tools work best where (or at what stage of training) needs to be carried out.

5.4 Concluding Remarks

An increasing emphasis on good quality training has occurred over the last decade. There is an expanding appreciation for all involved in training of the importance of providing useful feedback and ensuring that training opportunities are not wasted. Much of the work within this thesis serves to demonstrate that excellent training does occur, and can be achieved.

Greater appreciation of those that train is required in order to provide them with adequate time to ensure a high quality training experience. There have previously not been standards for training, but the GMC has advocated the recognition and approval of trainers[88]. This needs to be accompanied by a structured and robust method of assessing training from undergraduate to postgraduate years.

Whilst evidence based medicine has become the gold-standard for patient care, future interventions in training should be held to the same level of scrutiny to ensure the highest standards are achieved.

References:

1. Phillips AW, Madhavan A. A critical evaluation of the intercollegiate surgical curriculum and comparison with its predecessor the “Calman” curriculum. *J Surg Educ*.;70:557–62.

2. Phillips AW, Dent B, Navidi M, et al. Trainee Involvement in Ivor Lewis Esophagectomy Does Not Negatively Impact Outcomes. *Ann Surg*. 2018;267:94–98.

3. Phillips A, Jones A. Workplace based assessments in surgical training: Valid and reliable? *Bull R Coll Surg Engl*. 2015;97:e19–e23.

4. Nesbitt CI, Phillips AW, Searle RF, et al. Randomized trial to assess the effect of supervised and unsupervised video feedback on teaching practical skills. *J Surg Educ*.;72:697–703.

5. Phillips AW, Matthan J, Bookless LR, et al. Individualised Expert Feedback is Not Essential for Improving Basic Clinical Skills Performance in Novice Learners: A Randomized Trial. *J Surg Educ*. 2017;74:612–620.

6. Phillips AW, Madhavan A, Bookless LR, et al. Surgical Trainers’ Experience and Perspectives on Workplace-Based Assessments. *J Surg Educ*. 2015;72:979–984.

7. Phillips A, Lim J, Madhavan A, et al. Case-based discussions: UK surgical trainee perceptions. *Clin Teach*. 2016;13:207–212.

8. Nesbitt C, Phillips AW, Searle R, et al. Student Views on the Use of 2 Styles of Video-Enhanced Feedback Compared to Standard Lecture Feedback During Clinical Skills Training. *J Surg Educ*. 2015;72:969–973.

9. Kotsis S V, Chung KC. Application of the "see one, do one, teach one" concept in surgical training. *Plast Reconstr Surg*. 2013;131:1194–201.

10. Qayumi K, Vancouver. Surgical Skills Lab: A Hub for Competency Training. *J Investig Surg*. 2010;23:48–56.

11. Vozenilek J, Huff JS, Reznek M, et al. See One, Do One, Teach One: Advanced

Technology in Medical Education. *Acad Emerg Med*. 2004;11:1149–1154.

12. Rohrich RJ. “See one, do one, teach one”: An old adage with a new twist. *Plastic and Reconstructive Surgery*, 2006, 257–258.

13. Greensmith M, Cho J, Hargest R. Changes in surgical training opportunities in Britain and South Africa. *Int J Surg*. 2016;25:76–81.

14. Carlin AM, Gasevic E, Shepard AD. Effect of the 80-hour work week on resident operative experience in general surgery. *Am J Surg*. 2007;193:326–330.

15. Maxwell AJ, Crocker M, Jones TL, et al. Implementation of the European Working Time Directive in neurosurgery reduces continuity of care and training opportunities. *Acta Neurochir (Wien)*. 2010;152:1207–10.

16. Lewis FR, Klingensmith ME. Issues in General Surgery Residency Training—2012. *Ann Surg*. 2012;256:553–559.

17. Gough IR. The impact of reduced working hours on surgical training in Australia and New Zealand. *Surgeon*. 2011;9 Suppl 1:S8-9.

18. Antiel RM, Reed DA, Van Arendonk KJ, et al. Effects of duty hour restrictions on core competencies, education, quality of life, and burnout among general surgery interns. *JAMA Surg*. 2013;148:448–55.

19. Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon Volume and Operative Mortality in the United States. *N Engl J Med*. 2003;349:2117–2127.

20. Birkmeyer JD, Siewers AE, Finlayson EVA, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128–37.

21. Fitzgerald JEF, Milburn JA, Khera G, et al. Clinical Fellowships in Surgical Training: Analysis of a National Pan-specialty Workforce Survey. *World J Surg*. 2013;37:945–952.

22. Calman K. Hospital doctors: training for the future. *Br J Obstet Gynaecol*. 1995;102:354–6.

23. Taylor RM. HC 25-I Modernising Medical Careers Third Report of Session 2007–08 Volume I Report, together with formal minutes. *Labour*. Available from:

www.parliament.uk/healthcom. 2008. Accessed March 28, 2017.

24. Thomas C, Griffiths G, Abdelrahman T, et al. Does UK surgical training provide enough experience to meet today's training requirements. *BMJ Careers*.;May Available from: http://careers.bmj.com/careers/advice/. 2015.

25. Donaldson L. Unfinished business: proposals for the reform of the senior house officer grade- a paper for consultaton. 2002.

26. Modernising Medical Careers The next steps The future shape of Foundation, Specialist and General Practice Training Programmes Available from: http://webarchive.nationalarchives.gov.uk/20130107105354/http:/dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4079532.pdf. 2004. Accessed March 28, 2017.

27. The New Deal: Plan for action. The Report of the Working Group on Specialist Training. Leeds; 1994.

28. Training for the future, The Report of the Working Group on Specialist Medical Training. London; 1993.

29. Epstein J. Reduced working hours and surgical training in the UK. *Surgery*. 2004;22:i–ii.

30. Medicine Available from: http://www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-Medicine.pdf. 2003.

31. Health D of. Up to 1,500 extra medical training places announced - GOV.UK Available from: https://www.gov.uk/government/news/up-to-1500-extra-medical-training-places-announced. Accessed March 28, 2017.

32. Hanly D. Report of the National Task Force on Medical staffing Available from: http://www.dohc.ie/publications/hanly_report.html. 2003.

33. Spencer J. Learning and teaching in the clinical environment. *BMJ*. 2003;326:591–4.

34. Busari JO, Weggelaar NM, Knottnerus AC, et al. How medical residents perceive the quality of supervision provided by attending doctors in the clinical setting. *Med Educ*. 2005;39:696–703.

35. Batchelder AJ, Rodrigues CMC, Lin L-Y, et al. The role of students as teachers: four years' experience of a large-scale, peer-led programme. *Med Teach*. 2010;32:547–51.

36. Hampton T. US medical school enrollment rising, but residency programs too limited. *JAMA*. 2008;299:2846.

37. E. Preston-Whyte, R. Clark, S. Pete M. The views of academic and clinical teachers in Leicester Medical School on criteria to assess teaching competence in the small-group setting. *Med Teach*. 1999;21:500–505.

38. Consultants' Contract: Annual Appraisal for Consultants. London; 2003.

39. Tomorrow's Doctors Available from: http://www.gmc-uk.org/med_ed/tomdoc.htm. 2006.

40. Calman KC. Medical education--a look into the future. *Postgrad Med J*. 1993;69 Suppl 2:S3-5.

41. Whitcomb ME, Vice S. Educating Doctors to Provide High Quality Medical Care A Vision for Medical Education in the United States Available from: www.aamc.org. 2005. Accessed March 26, 2017.

42. Grant J. Service-based learning in hospital medicine: integrating patient care and training in the early postgraduate years. In: Jolly B, Rees L, eds. Medical Education in the Millenium. Oxford: Oxford University Press; 1998:156–169.

43. Stanley P. Structuring ward rounds for learning: can opportunities be created? *Med Educ*. 1998;32:239–43.

44. Kolb D. Experiential learning: Experience as the source of learning and development. Englewood Cliffs, NJ.: Prentice Hall; 1984.

45. Campbell C, Parboosingh J, Gondocz T, et al. A study of the factors that influence physicians' commitments to change their practices using learning diaries. *Acad Med*. 1999;74:S34-6.

46. Schon D. Educating the Reflective Practitioner. Josse-Bass; 1987.

47. Papastavrou E, Dimitriadou M, Tsangari H, et al. Nursing students' satisfaction of the clinical learning environment: a research study. *BMC Nurs*. 2016;15:44.

48. Bowrey DJ, Kidd JM. How Do Early Emotional Experiences in the Operating Theatre Influence Medical Student Learning in This Environment? *Teach Learn Med*. 2014;26:113–120.

49. Parsell GJ, Bligh J. The changing context of undergraduate medical education. *Postgrad Med J*. 1995;71:397–403.

50. WFME. WFME Task Force on Defining International Standards in Basic Medical Education. Report of the Working Party. *Med Educ*. 1999;34:665–675.

51. Bastian H, Glasziou P, Chalmers I. Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up? *PLoS Med*. 2010;7:e1000326.

52. Firth-Cozens J. New stressors, new remedies. *Occup Med (Lond)*. 2000;50:199–201.

53. Firth-Cozens J, Moss F, Rayner C, et al. The effect of 1-year rotations on stress in preregistration house officers. *Hosp Med*. 2000;61:859–60.

54. Griffith CH, Wilson JF, Haist SA, et al. Relationships of how well attending physicians teach to their students' performances and residency choices. *Acad Med*. 1997;72:S118-20.

55. Assessment working Group. Developing and maintaining an assessment system- a PMETB guide to good practice. 2007.

56. Sadler R. Formative assessment and the design of instructional systems. *Instr Sci*. 1989;119–144.

57. Branch WT, Paranjape A. Feedback and reflection: teaching methods for clinical settings. *Acad Med*. 2002;77:1185–8.

58. Gipps C. Socio-Cultural aspect of assessment. *Rev Educ Res*.;24:355–392.

59. Shepard L. The role of assessment in a learning culture. *Educ Res*. 2000;4–14.

60. Anziani H, Durham J, Moore U. The relationship between formative and summative assessment of undergraduates in oral surgery. *Eur J Dent Educ*. 2008;12:233–238.

61. Hattie J, Timperley H. The Power of Feedback. *Rev Educ Res*. 2007;77:81–112.

62. Veloski J, Boex JR, Grasberger MJ, et al. Systematic review of the literature on

assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach*. 2006;28:117–128.

63. Burch VC, Seggie JL, Gary NE. Formative assessment promotes learning in undergraduate clinical clerkships. *S Afr Med J*. 2006;96:430–3.

64. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65:S63-7.

65. Torsney KM, Cocker DM, Slesser AAP. The Modern Surgeon and Competency Assessment: Are the Workplace-Based Assessments Evidence-Based? *World J Surg*. 2015;39:623–633.

66. Phillips AW, Bookless LR. In Response to: Torsney KM, Cocker DM, Slesser AAP (2015) The Modern Surgeon and Competency Assessment: Are the Workplace-Based Assessments Evidence-Based? *World J Surg*. 2015;39:2838–2839.

67. Boulet JR, McKinley DW, Norcini JJ, et al. Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Adv Health Sci Educ Theory Pract*. 2002;7:85–97.

68. Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Ann Intern Med*. 2004;140:874–81.

69. Aggarwal R, Balasundaram I, Darzi A. Training opportunities and the role of virtual reality simulation in acquisition of basic laparoscopic skills. *J Surg Res*. 2008;145:80–6.

70. Al-Elq AH. Simulation-based medical teaching and learning. *J Family Community Med*. 2010;17:35–40.

71. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. *N Engl J Med*. 2006;355:2664–9.

72. Sarker SK, Patel B. Simulation and surgical training. *Int J Clin Pract*. 2007;61:2120–2125.

73. McCloy R. Virtual Reality in Surgery. *BMJ*. 2001;323:912.

74. Kapadia MR, DaRosa DA, MacRae HM, et al. Current assessment and future directions of surgical skills laboratories. *J Surg Educ*. 2007;64:260–5.

75. Walker M, Peyton J. Teaching in Theatre. In: Peyton J, ed. Teaching and learning in medical practice. Manticore Europe Ltd; 1998:171–180.

76. Fitts P, Posner M. Learning and skilled performance. In: Human Perfomance. Belmont CA: Brock-Cole; 1967.

77. Stefanidis D, Scerbo MW, Sechrist C, et al. Do novices display automaticity during simulator training? *Am J Surg*. 2008;195:210–3.

78. Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. *J Surg Educ*.;65:77–83.

79. Moulton C-AE, Dubrowski A, Macrae H, et al. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Ann Surg*. 2006;244:400–9.

80. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. 2002;236:458-63–4.

81. Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg*. 2004;91:146–150.

82. National Student Survey: Findings and trends 2006-2010 Available from: www.hefce.ac.uk/media/hefce/content/pubs/2011/201111/11_11.pdf. Accessed May 24, 2012.

83. Asch DA, Nicholson S, Srinivas S, et al. Evaluating Obstetrical Residency Programs Using Patient Outcomes. *JAMA*. 2009;302:1277.

84. Pereira EAC, Dean BJF. British surgeons' experiences of mandatory online workplace-based assessment. *J R Soc Med*. 2009;102:287–93.

85. Pereira EAC, Dean BJF. British surgeons' experiences of a mandatory online workplace based assessment portfolio resurveyed three years on. *J Surg Educ*.;70:59–67.

86. Ali J, Goh A. Student perceptions of workplace-based assessment. *Clin Teach*.

2017;14:319–324.

87. de Jonge LPJWM, Timmerman AA, Govaerts MJB, et al. Stakeholder perspectives on workplace-based performance assessment: towards a better understanding of assessor behaviour. *Adv Heal Sci Educ*. 2017;22:1213–1243.

88. Recognititon and approval of trainers Available from: http://www.gmc-uk.org/education/10264.asp.

89. Stenbacka C. Qualitative research requires quality concepts of its own. *Manag Decis*. 2001;39:551–555.

90. Lincolon Y, Guba E. Naturilistic Inquiry. Los Angeles: Sage Publications; 1985.

91. Oppenheim A. Questionnaire Design, Interviewing and Attitude Measurement. London: Pinter Publishers Ltd; 1992.

92. Silverman D. Interpreting Qualitative Data. London: Sage Publications; 1993.

93. Cohen L, Manion L, Morrison K. Research Methods in Education. 5th ed. London: Routledge Falmer; 2005.

94. Strauss A, Corbin J. Basics of qualitative research: Techniques and procedures for developing grounded theory. 2nd Editio. London: Sage Publications; 1998.

95. ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ*. 2005;39:1176–1177.

96. Black D. An end to box ticking: an overhaul of competency based education. *BMJ Careers*.

97. Sterkenburg A, Barach P, Kalkman C, et al. When Do Supervising Physicians Decide to Entrust Residents With Unsupervised Tasks? *Acad Med*. 2010;85:1408–1417.

98. Ten Cate O, Chen HC, Hoff RG, et al. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach*. 2015;37:983–1002.

99. Council GM. Tomorrow’s doctors: Duties of a doctor registered with the General Medical Council. 2009.

100. Watling CJ, Lingard L. Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process. *Adv Health Sci Educ Theory Pract*. 2012;17:183–94.

101. Cherry-Bukowiec JR, Machado-Aranda D, To K, et al. Improvement in acute care surgery medical student education and clerkships: use of feedback and loop closure. *J Surg Res*. 2015;199:15–22.

102. UK review of the provision of information about higher education: National Student Survey results and trend analysis 2005-2013 Available from: http://www.thestudentsurvey.com/. 2013.

103. Sargeant J, Mann K, Sinclair D, et al. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Adv Health Sci Educ Theory Pract*. 2008;13:275–88.

104. Andreatta PB, Hillard M, Krain LP. The impact of stress factors in simulation-based laparoscopic training. *Surgery*. 2010;147:631–9.

105. Maher Z, Milner R, Cripe J, et al. Stress training for the surgical resident. *Am J Surg*. 2013;205:169–174.

106. Müller MP, Hänsel M, Fichtner A, et al. Excellence in performance and stress reduction during two different full scale simulator training courses: a pilot study. *Resuscitation*. 2009;80:919–24.

107. Arora S, Aggarwal R, Moran A, et al. Mental practice: effective stress management training for novice surgeons. *J Am Coll Surg*. 2011;212:225–33.

108. Wetzel CM, George A, Hanna GB, et al. Stress Management Training for Surgeons—A Randomized, Controlled, Intervention Study. *Ann Surg*. 2011;253:488–494.

109. Tooke J. Aspiringto Excellence Available from: http://www.medschools.ac.uk/AboutUs/Projects/Documents/Final MMC Inquiry Jan2008.pdf.

110. Temple J. "Time for training": A review of the impact of the European Working Time Directive on the quality of Training. 2010.