



Assessment of surgical competence

A Darzi and S Mackay

Qual. Health Care 2001;10;ii64-ii69
doi:10.1136/qhc.0100064..

Updated information and services can be found at:

http://qshc.bmj.com/cgi/content/full/10/suppl_2/ii64

These include:

References

This article cites 21 articles, 1 of which can be accessed free at:

http://qshc.bmj.com/cgi/content/full/10/suppl_2/ii64#BIBL

1 online articles that cite this article can be accessed at:

http://qshc.bmj.com/cgi/content/full/10/suppl_2/ii64#otherarticles

Rapid responses

You can respond to this article at:

http://qshc.bmj.com/cgi/eletter-submit/10/suppl_2/ii64

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

To order reprints of this article go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to *Quality in Health Care* go to:

<http://journals.bmj.com/subscriptions/>

Assessment of surgical competence

A Darzi, S Mackay

Abstract

This paper examines the issues that arise in the broad area of competence assessment in surgical practice, with particular reference to the objective assessment of technical skill which has historically been the weakest aspect of assessment in surgical training. To facilitate a thorough appraisal of competence, a simple model of surgical practice is advanced, followed by a review of both current and experimental methods of assessing technical skill. The review comprises not only the published literature, but also work (both from the authors' and other groups) that is in progress or under consideration for publication. Significant issues in the implementation of these new technologies, especially the necessary further validation, and the imperative to demonstrate that the process introduced does indeed improve the outcomes are discussed.

(Quality in Health Care 2001;10(Suppl II):ii64–ii69)

Keywords: competence assessment; surgery; learning

In recent years there has been an increasing debate about competence in medical practice, perhaps especially so in surgery.¹ To be effective practitioners, all doctors need a wide range of skills. These include, in addition to technical competencies, the ability to listen and the ability to inform and to guide people as they make choices about their treatment. Care is invariably the result of a team working together; doctors must be effective team members and, at times, team leaders. Moreover, as we are learning from the analysis of circumstances where things go horribly wrong, problems with care are often linked to organisational and system failure.² Hence senior doctors—the leaders—need the organisational skills to manage change and to motivate teams. Finally, those who supervise doctors in training need the skills to train and to assess trainees.

At the centre of surgical practice, however, are the surgeon's technical skills. There have been instances in which it has been suggested that poor outcomes were the result of inadequate technical performance. While acknowledging that other skills are important for good practice, it is crucial that surgeons, and others who undertake technical tasks, are technically competent and that the methods for assessing those competencies are robust. Concerns about the assessment of technical competencies (which has historically been difficult) have given rise to an increasing interest in the objective scientific measurement of technical performance.^{1 3}

Technologies are being developed that will allow significant developments in this area.⁴ However, even if we limit discussion of the surgical skills and competencies to the defining clinical and practical ones, it becomes clear that the necessary skills remain multiple and complex. For example, advanced technological assessment of practical skills can have little meaning unless it is clear that the surgeon also knows how and when to use those skills.

We have developed technological approaches to the assessment of surgical skills, using a model of the process of clinical care that outlines its components to shape a profile of the individual competencies that contribute to overall competence. These competencies are described and the methods of assessment, from the traditional to those being developed, are considered. Where appropriate, comments are made concerning the usefulness of a given methodology. The current state of technical skills assessment and the issues that arise in trying to reform the process of assessment are discussed.

Model of care

When considering the objective assessment of surgical competence it is useful to have a framework for categorising the important aspects of the process of surgical care. This promotes a more complete approach to assessment and reduces the possibility of "gaps" developing. In broad terms the process of surgical care has four components: diagnosis, plan of treatment, technical performance, and post-operative care.

DIAGNOSTIC ABILITY

This is essential for all areas of clinical medicine. Careful history taking, the ability to listen carefully to the patient's own story, the ability to perform a physical examination and

Key messages

- There is a growing interest in "competence assessment", a process whereby all aspects of a doctor's practice are formally validated by an objective assessment process.
- In surgery the area of technical performance has historically been most problematic in terms of objective assessment.
- New technologies have been developed in the past decade that seem likely to facilitate objective assessment of technical skill.
- Significant work remains to integrate these processes into training programmes and to demonstrate an improvement in outcomes.

Imperial College of Science Technology and Medicine, Faculty of Medicine, Academic Surgical Unit, St Mary's Hospital, London W2 1NY, UK
A Darzi, professor of surgery and head of department
S Mackay, honorary research fellow

Correspondence to:
Mr S Mackay
s.mackay@ic.ac.uk

An 82 year old woman complaining of severe abdominal pain present for 90 minutes is brought to casualty by ambulance. She has been in good health but has atrial fibrillation for which she takes digoxin and has had congestive cardiac failure for which she takes frusemide. She has no past surgical history and is afebrile. Her pulse is 90, in atrial fibrillation, and blood pressure is 110/65. Abdominal examination reveals severe generalised tenderness but no guarding. There is no evidence of any hernias. No abdominal aneurysm is palpable and the circulation in her legs is normal. The erect chest radiograph is normal, and supine and erect plain abdominal radiographs suggest small bowel obstruction.

Box 1 Example of symptoms for establishing a diagnosis.

to explain, recommend and interpret diagnostic tests are essential tools for all doctors. The essential information, once gathered, enables a diagnosis or differential diagnoses to be formulated. A surgeon asked to see a patient with abdominal pain will use these skills to make a diagnosis. In the example given in box 1 various aspects of the presentation—the patient's age, the presence of an irregular heart beat, and the results of the tests—mean that the surgeon will be particularly concerned that the patient's symptoms are caused by "mesenteric ischaemia" (failing blood supply to the intestine which can lead to gangrene of the bowel and death), although other possible diagnoses include perforation of the intestine or other viscus or pancreatitis (inflammation of the pancreas). Assessment should include some way of indicating just how well a trainee can use the information available and work out the probable and possible causes of a patient's problem.

TREATMENT PLAN

Treatment plan(s) can be recommended once a diagnosis is established. This involves consideration of the available options, evaluation of the strengths and weaknesses of each option, and discussion with the patient. A treatment plan may include a surgical intervention, but not always. A surgeon has to be prepared to change plan if, for example, further information becomes available. Thus two components of care—the diagnostic process and establishment of a treatment plan—are not always separate and may intertwine, especially if the problem is complex.

In the example in box 1, if the patient had mesenteric ischaemia she would require an emergency operation but, before proceeding, the surgeon would need to rule out the possibility of pancreatitis (which does not usually require surgery) and would wait for analysis of the blood amylase while the patient was being prepared for anaesthesia and then surgery. If the blood amylase result indicated the probability of pancreatitis, the operation could be cancelled and further investigations

requested relevant to the investigation of pancreatitis.

Assessment must therefore examine not only a trainee's ability in diagnosis and treatment as isolated components of the clinical process, but also how well he or she can integrate these aspects in the complex and relatively uncertain arena of clinical practice.

TECHNICAL PERFORMANCE

When considering technical performance it is tempting to concentrate solely on technical dexterity but there are, in fact, other essential aspects of competence relevant to technical performance.

Firstly, there is the surgeon's *judgement*. This term is widely and loosely used, often to refer to a skill that broadly equates with an amalgam of "diagnostic ability" and "treatment plan". In this context, however, we are using it very specifically to refer to the decision making that takes place during a surgical (or other) procedure. In the case described in box 1, after considering all the information the patient is thought to have mesenteric ischaemia and undergoes an emergency laparotomy at which the diagnosis is confirmed. Achieving the diagnosis and establishing a treatment plan have required one set of skills. Once the operation has started the surgeon has to make a series of important decisions within a relatively constrained time frame. It is this decision making process that is referred to here as judgement. In this case the surgeon has to decide in the operating theatre whether a given loop of bowel is irrevocably damaged by poor blood supply and must be removed or whether it is likely to survive. This decision may be the most important part of the procedure; if bowel which is too damaged is left behind to become gangrenous the patient will deteriorate and die but, likewise, overenthusiastic resection (removal) of questionable bowel may leave the patient with lifelong nutritional problems.

The second aspect of technical skill is *knowledge*. This refers to the knowledge base required to implement the decisions made as part of a surgeon's judgement. In the example presented in box 1, if a loop of bowel was not dusky but rather black and clearly necrotic, then it does not take a surgeon to decide that a resection is necessary—it would be expected that any medical graduate would recognise this characteristic pathology if it were on view. However, having decided that a resection is necessary, the question arises of how to perform the intestinal resection. Having this information (for all the procedures in his/her repertoire) is a key aspect of technical competence for a practising surgeon.

The third aspect of technical skill is *dexterity*. In the simplest terms this refers to the pure psychomotor aspects of the task at hand—that is, the dexterity required to execute the planned procedure. This is more than being able to demonstrate quick fluent movements that may look impressive to an observer. It includes, for example, being able to suture tissue accurately and tie knots that are just tight enough to promote healing (are functional and

prevent fluid leaking) but are not so tight as to cause tissue damage. Another important aspect of surgical dexterity is how a surgeon actually handles the tissues—doing this expertly will minimise trauma and speed healing and recovery.

Assessment of technical performance therefore needs to include the range of competencies necessary for carrying out a procedure effectively. Again, assessment should include how a surgeon acts and reacts as a situation unfolds and, ideally, how he/she combines expert knowledge, judgement, and dexterity.

POSTOPERATIVE CARE

Postoperative care is the responsibility of the surgeon and is shared with others including the anaesthetic team and the nurses and physiotherapists working on the postoperative wards. Routine care pathways, although well rehearsed, may themselves be complex and challenging, especially after a major operation in a critically ill patient. The surgeon will need his/her diagnostic skills at this stage to detect any complications at an early stage. If a complication has been diagnosed, a treatment plan must be established to deal with it.

Hence, the model of care is a loop in which diagnosis is re-evaluated and, if indicated, further investigations are done as the care progresses. It is easy to see that several aspects of the care process may be happening simultaneously, especially in a complex case. Using this model allows the observer to categorise and appraise the range of skills needed for the surgeon to deliver competent care and to determine the method most appropriate for assessment of each aspect.

Methods of assessment

CONVENTIONAL WRITTEN EXAMINATIONS

Written examinations have long been a mainstay of medical examinations. Formats include essays, short answers, multiple choice questions and, more recently, extended matching questions. The advantages of written examinations are that both the questions and the marking scheme can be standardised and the process is easily demonstrated to be objective and fair. However, the content being assessed is limited either in scope (essay format) or depth (multiple choice format), and may not sufficiently assess the complex attributes essential for good practice.

The written examination is, in general, an effective approach to assessing factual knowledge but has limited application for assessment of decision making ability. Looking at the model of care, this format could most usefully be used to assess facts—for example, in the case described in box 1 this would include the anatomy of the arterial supply to the mesentery, knowledge that is necessary for undertaking bowel resection—and perhaps is least useful for assessing some of the complexities of the diagnostic process or capacity for intraoperative judgement.

CONVENTIONAL VIVA VOCE EXAMINATIONS

The “viva” has been an established part of the process of medical assessment for years. This examination allows some assessment of facts but may also allow exploration of the integration of information—for example, in a traditional “long case” examination where a candidate is assessed around a diagnostic consultation with a patient. The strength of this format is that the examiners may explore topics with the candidate in greater or lesser depth, as appropriate, and may also examine the process by which a candidate comes to a particular decision or view. The weakness is that the process is not standardised and is difficult to make objective, so it may be unfair to some. Furthermore, the *viva voce* is a potentially threatening process and some candidates may be disadvantaged by being more intimidated than others.

Such examinations are most useful for exploring how a candidate arrives at a diagnosis and for assessing the judgement necessary in intraoperative decision making. The assessors can use the situation to re-state questions with subtle variations to ensure that the candidate truly understands the rationale for the decisions that he/she proposes.

OBJECTIVE STRUCTURED CLINICAL EXAMINATIONS

Objective structured clinical examinations (OSCEs) are based on a series of stations, each of which has a self-contained question/item.⁵⁻⁷ The format is typically of candidates rotating through several stations that can be taken in any order in a “round robin” manner. The potential content is wide ranging and, importantly, can be standardised. Questions/items may include, for example, pathology results with specific questions, radiographs for interpretation, discussing a diagnosis, taking a history from an actor/patient, or evaluation of clinical scenarios. Marking is by trained observers placed at each station. Candidates at each station are marked by the same observer in a standardised marking process which uses a scoring system that is established according to agreed objective criteria.

The great advantage of this process is that a wide variety of material may be examined in a highly standardised way. The disadvantage is that the depth of assessment is often limited by time constraints. OSCEs are of great value in assessing the interpretation of information—whereas a written examination can ask about a test, the OSCE allows the presentation of a set of results and seeks an appraisal by the candidate. However, these assessments are limited in their capacity to explore a candidate’s understanding of complex issues; the cases and situations presented may be suitable for in depth analysis but the time constraints and the requirements of a standardised marking schedule make this difficult.

OBJECTIVE STRUCTURED ASSESSMENTS OF TECHNICAL SKILLS

Objective structured assessments of technical skills (OSATS) is a methodology based, to some extent, on the OSCE concept that was

developed by Reznick and co-workers in Toronto.⁸⁻¹⁰ In this form of assessment the candidate performs a standardised surgical task while being observed by at least two assessors. The tasks studied in this assessment have included: placing sutures in a pad of synthetic skin; joining together two cut ends of bowel; and inserting an intercostal catheter (a tube placed into the chest between the ribs to drain air or fluid from the lung). The observers mark the performance of the task using two marking systems—a checklist and a global scoring sheet. The checklist comprises a series of yes/no items which have been developed by analysis of the task and also of the specific tuition that has been provided in skills training sessions. The global scoring sheet comprises eight items, each of which is marked from 1 to 5. The items assessed include tissue handling skills, flow of operation, and familiarity with the technique. Examples of poor (score 1), average (score 3), and excellent performance (score 5) are given as guidelines for the observers.

Both global and checklist scoring systems have been validated.⁹ In general, global scoring assesses generic aspects of technical performance and has a broad applicability whereas checklists are task specific. A new checklist must be developed and validated for each new task included in the assessment. Experience suggests that global scoring is a more effective discriminator between subjects than the checklist,¹⁰ perhaps because the checklist items, of necessity, have to be relatively straightforward components of a procedure and should not call upon the observer to exercise significant judgement in deciding whether to mark “yes” or “no”.

OSATS have been widely used by Reznick *et al* and are being increasingly used elsewhere. They are useful in assessing technical skills in terms of knowledge and dexterity aspects but they do not offer the scope to assess judgement as the tasks are highly standardised. Currently, the methodology is well established as a research tool and is moving towards implementation within training schemes in some countries.

HAND-MOTION ANALYSIS

OSATS represents a step forward in assessing dexterity, but there is certainly room for other techniques. Our group has developed a device (Imperial College Surgical Assessment Device, ICSAD) that uses motion analysis to determine how many movements a subject uses to perform a standardised surgical task.^{11 12} The motion analysis currently uses an alternating current (AC) electromagnetic (EM) system (although it is equally applicable to direct current EM systems or to ultrasound or infrared based systems) in which passive trackers are attached to the dorsum of each hand. When the hands are moved within the magnetic field generated by the system a current is induced in the trackers and the analysis of this current allows the tracking device to determine the position in space of that tracker. These positional data are currently measured at 20 Hz but measurements at more than 100 Hz

are possible. The ICSAD system comprises software that takes the raw positional data and converts them to information on the number of movements and the path length—as well as the process of integrating the three dimensional coordinates, various filters are applied to minimise noise in the measurement.

The measures have been shown to be an effective index of technical skill in both laparoscopic¹¹ and open¹² procedures, and demonstrate good concordance with OSATS scoring. Surgeons of varying levels of experience each performed two tasks (intestinal suturing and vascular suturing) and were assessed by ICSAD and OSATS.¹³ Both methods of measurement showed a significant relation between experience and performance, and there was good correlation between the two measures for each task (this work has been submitted for publication but is not yet in press).

Given that the technique relies on a standardised task and technique, it is most applicable in assessing the knowledge and dexterity components of technical performance as the standardisation removes the opportunity for the subject to display judgement.

VIRTUAL REALITY

Virtual reality (VR) is a technology that holds out the exciting prospect of including simulation as part of the training and assessment of surgical performance.^{14 15} It offers the opportunity to learn a new skill without the pressure of a clinical situation. Moreover, it is theoretically possible for the learner to have repeated practice and tuition on any weak aspects of a given procedure. Beyond this, VR offers very detailed feedback on the progress that is being made and may allow more subtle measurement of performance than is possible in a “real world” setting. From the patient’s point of view, it is obviously preferable that trainees’ attempts at a new procedure are performed on a simulator than on a patient.

The Minimally Invasive Surgical Trainer—Virtual Reality (MIST VR; Mentice, Gothenburg, Sweden) was the first virtual reality system to be examined in detail in the area of medical skills assessment.¹⁶⁻²⁰ Studies to date have confirmed that it has validity as an assessment tool but not as a training tool. Thus, performance on the device correlates with laparoscopic ability¹⁸ but it has not been possible to show that training in VR supplants or reduces the need for training in the clinical environment. MIST VR is a “low fidelity” system which attempts to replicate the skills of laparoscopic operating but not the appearance (the virtual environment consists of a wire cage in which various geometric objects may be manipulated). It therefore has no applicability in the area of knowledge or judgement (as no operation is simulated), although it appears to be effective in assessing dexterity. There are a series of “high fidelity” simulators under assessment and it seems likely that these will also allow assessment of candidates’ knowledge of the steps in a procedure. Some of the early results in this area have been disappointing,²¹

but subsequent experience suggests that this may be due to overambitious goals.

At this stage VR must be seen as experimental technology. It is true that there are certain procedures such as endoscopy for which current computing technology is eminently suitable and for which high fidelity simulators are available. However, there are, as yet, few studies validating these devices as assessment tools and none validating them as training tools. There is no doubt that, as VR technology develops, simulators are likely to have an increasing role in future approaches to training and assessment.

COMPETENCE DAY

The new technologies discussed above hold significant potential for objective assessment of technical performance. Each of the assessment techniques has its strengths and weaknesses, and it may be that certain subjects will also find one method more intimidating or difficult than another (and hence underperform). Based on these techniques and the model of competence proposed above, we have recently validated a six task “competence day” assessment for senior house officers at the level of the Membership of the Royal Colleges of Surgeons (MRCS). The underlying premise is that the assessment will be more robust if candidates are assessed on multiple parameters using a variety of measures. In this way, if a candidate suffered a relative disadvantage on one method of testing, the effect would not be active across the whole assessment. The six tasks validated comprise one each of OSCE and VR, and two each of OSATS and ICSAD.²² The tasks assessed in the initial study comprised: a conventional OSCE station to test knowledge of sutures, instruments and surgical equipment; suturing a synthetic skin pad (ICSAD); tying a surgical knot (ICSAD); suturing synthetic small intestine (OSATS); excising a lesion in synthetic skin (OSATS); and the MIST VR. Overall reliability (Cronbach’s alpha) of this six part examination, which actually included 19 separate analyses, was 0.71 (this work has been submitted for publication but is not yet in press).

These results are promising and the underlying hypothesis has good face validity, but more work is necessary before such a process can be introduced within the training and assessment processes. One particular issue is whether it is reasonable to conceptualise this competence day as being like a driving test or whether it should be modelled on the conventional undergraduate/postgraduate examination process. The distinction is that the content of the driving test is (essentially) known to the subject before he/she embarks upon it and the aim is to see that a necessary level is achieved in each of several fundamental skills, whereas the conventional examination is predicated upon the idea that the content must be kept secret as it comprises only a snapshot of the knowledge being tested and will be invalidated if the subjects have any prior idea of the content. It is

also true that the driving test offers the opportunity for an instant fail whereas the conventional examination does not have such “killer items” and a subject fails only upon the basis of an inadequate aggregate mark. We suggest that a process like the driving test is most reasonable for junior trainees (smaller range of fundamental skills, hence the process is more comprehensive), whereas the examination of senior trainees or the revalidation of practising surgeons will require a broader series of assessments.

Discussion

Genuine and reasonable concerns over technical competence have driven the development of a new area of research—the objective assessment of technical skill. Traditionally, this has been the area of surgical competence that is least well assessed; however, this is set to change with the inevitable implementation of new methodologies such as OSATS, ICSAD, and VR systems, as well as a change in thinking within the surgical community which now regards this as a priority. The competence day approach represents a further step along this pathway and has been developed to overcome some of the potential shortcomings of using just one new technology which has been designed as an integrated skills examination aimed at trainees at a certain level.

However, technical performance cannot be seen as an isolated end in itself and the drive for effective objective assessment is part of a growing “competence movement” within the profession. In this context, there is value in using a simple but inclusive model of overall competence to ensure that any new process is thorough and complete.

There are several groups working in this area in the UK and around the world. Many are based within surgical colleges and have the remit to integrate assessment and formal training into the hitherto hospital based training. In this sense, Professor Reznick’s group are the most advanced,^{23–25} working in a setting where the surgical trainees are required to attend the skills laboratory (and are given protected time to attend) for skills training and assessment. In Canada postgraduate training is the responsibility of the universities and hence there is a seamless progression from medical student to resident to advanced trainee. In the UK several groups are involved in postgraduate training (Royal Colleges, Deaneries, NHS, and GMC) and no one group is currently in a position to develop and implement a thorough process that is applied across the country in all settings and across all levels of training. There is no doubt that this fact is a potential bar to the early and effective introduction of a skills assessment programme.

Technical skills are the least well assessed component of the clinical process because assessment techniques currently in use are highly subjective and are poorly standardised and validated. The present assessment system relies on retrospective reporting from the surgeons for whom a given trainee has worked

over the preceding months. The raters (surgeons) have no formal training in this task and very little guidance in the area of setting standards for acceptable ability. The process is open to bias, especially as it is difficult to separate a subject's technical performance from an overall impression formed on the basis of other factors (punctuality, dedication, theoretical knowledge, etc). It seems clear that improvements in this process will come from implementation of the new approaches detailed above rather than some elaboration of the current process.

However, it should not be assumed that all the necessary work has been done and that there is an "off the shelf" technology that is just waiting to be put into practice.⁴ The results to date are promising but most studies have dealt with the validation of the measures examined by comparison with the seniority of the subjects. If formal assessment of technical skills is to be introduced, it will be necessary to demonstrate some benefit in terms of either individual outcome (such as detection of dangerously bad performance) or overall outcome (demonstrate that feedback and directed training will improve learning). This work has not been performed to date but it certainly should be afforded a high priority.

This idea of scrutinising the assessment process before and during implementation is an important one. Results in a skills laboratory may not reflect results in the operating theatre, and this is an issue that warrants careful consideration. The ideal would be to assess all aspects of competence in a realistic setting, or even in a true clinical setting. With this eventual aim, our group has been developing a "black box" recording system for an operating room or other complex medical environment. The system, the most advanced of its kind, aims to gather all available data and to store them in a manner whereby they may be interrogated efficiently for training, quality assurance, or research. It is hoped that such a system will make it possible to appraise any aspect of care (whether provided by the individual or the overall team) as delivered in a clinical setting. This is important given that current assessments (in all areas) tend to concentrate on the matter at hand which means that the subject is aware of the area of scrutiny as well as concentrating on the fact of being scrutinised. We feel that the ultimate aim for the "competence movement" should be effective, objective, and fair assessments carried out in a real or highly realistic setting.

- 1 Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. *BMJ* 1999;318:887-8.
- 2 www.bristol-inquiry.org.uk/final_report/index.htm.
- 3 Reznick RK. Teaching and testing technical skills. *Am J Surg* 1993;165:358-61.
- 4 Darzi A, Datta V, Mackay S. The challenge of objective assessment of surgical skill. *Am J Surg* 2001;181:484-6.
- 5 Cuschieri A, Gleeson FA, Harden RM, et al. A new approach to a final examination in surgery. Use of the objective structured clinical examination. *Ann R Coll Surg Engl* 1979;61:400-5.
- 6 Reznick RK, Smee S, Baumber JS, et al. Guidelines for estimating the real cost of an objective structured clinical examination. *Acad Med* 1993;68:513-7.
- 7 Sloan DA, Donnelly MB, Schwartz RW, et al. The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. *Ann Surg* 1995;222:735-42.
- 8 Faulkner H, Regehr G, Martin J, et al. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med* 1996;71:1363-5.
- 9 Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;84:273-8.
- 10 Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998;73:993-7.
- 11 Taffinder N, Smith SG, Huber J, et al. The effect of a second-generation 3D endoscope on the laparoscopic precision of novices and experienced surgeons. *Surg Endosc* 1999;13:1087-92.
- 12 Datta V, Mandalia M, Mackay S, et al. Relationship between skill and outcome in the laboratory based model. *Br J Surg* 2001;88(Suppl 1):46.
- 13 Datta V, Chang A, Mackay S, et al. The relationship between motion analysis and OSATS surgical assessments in the laboratory based model. Nashville: Association of Surgical Education, 2001.
- 14 Smith SG, Torkington J, Darzi A. Objective assessment of surgical dexterity using simulators. *Hosp Med* 1999;60:672-5.
- 15 Torkington J, Smith SG, Rees BI, et al. The role of simulation in surgical training. *Ann R Coll Surg Engl* 2000;82:88-94.
- 16 Wilson MS, Middlebrook A, Sutton C, et al. MIST VR: a virtual reality trainer for laparoscopic surgery assesses performance. *Ann R Coll Surg Engl* 1997;79:403-4.
- 17 Sutton C, McCloy R, Middlebrook A, et al. MIST VR. A laparoscopic surgery procedures trainer and evaluator. *Stud Health Technol Inform* 1997;39:598-607.
- 18 Taffinder N, Sutton C, Fishwick RJ, et al. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Stud Health Technol Inform* 1998;50:124-30.
- 19 Chaudhry A, Sutton C, Wood J, et al. Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface. *Ann R Coll Surg Engl* 1999;81:281-6.
- 20 Gallagher AG, McClure N, McGuigan J, et al. Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). *Endoscopy* 1999;31:310-3.
- 21 Smith S, Wan A, Taffinder N, et al. Early experience and validation work with Proceidicus VA—the Proslavia virtual reality shoulder arthroscopy trainer. *Stud Health Technol Inform* 1999;62:337-43.
- 22 Mackay S, Datta V, Morgan P, et al. Competence day: development of a panel of objective assessment tasks for senior house officers at MRCS level. *Br J Surg* 2001;88(Suppl 1):46.
- 23 MacRae H, Regehr G, Leadbetter W, et al. A comprehensive examination for senior surgical residents. *Am J Surg* 2000;179:190-3.
- 24 MacRae HM, Cohen R, Regehr G, et al. A new assessment tool: the patient assessment and management examination. *Surgery* 1997;122:335-43; discussion 343-4.
- 25 Reznick R, Regehr G, MacRae H, et al. Testing technical skill via an innovative "bench station" examination. *Am J Surg* 1997;173:226-30.