

Clinical review

Objective assessment of technical skills in surgery

Krishna Moorthy, Yaron Munz, Sudip K Sarker, Ara Darzi

In the past few years, considerable developments have been made in the objective assessment of technical proficiency of surgeons. Technical skills should be assessed during training, and various methods have been developed for this purpose

Department of
Surgical Oncology
and Technology,
Imperial College,
St Mary's Hospital,
London W2 1NY
Krishna Moorthy
clinical research fellow
Yaron Munz
clinical research fellow
Sudip Sarker
clinical research fellow
Ara Darzi
professor of surgery
Correspondence to:
K Moorthy
k.moorthy@
imperial.ac.uk

Surgical competence entails a combination of knowledge, technical skills, decision making, communication skills, and leadership skills. Of these, dexterity or technical proficiency is considered to be of paramount importance among surgical trainees. The assessment of technical skills during training has been considered to be a form of quality assurance for the future.¹ Typically surgical learning is based on an apprenticeship model. In this model the assessment of technical proficiency is the responsibility of the trainers. However, their assessment is largely subjective.² Objective assessment is essential because deficiencies in training and performance are difficult to correct without objective feedback.³

The introduction of the Calman system in the United Kingdom, the implementation of the European Working Time Directive, and the financial pressures to increase productivity⁴ have reduced the opportunity to learn surgical skills in the operating theatre. Studies have shown that these changes have resulted in nearly halving the surgical case load that trainees are exposed to.⁵ Surgical proficiency must therefore be acquired in less time, with the risk that some surgeons may not be sufficiently skilled at the completion of training.⁶ This and increasing attention of the public and media on the performance of doctors have given rise to an interest in the development of robust methods of assessment of technical skills.⁷ We review the research in this field in the past decade. Our objectives are to explore all the available methods, establish their validity and reliability, and examine the possibility of using these methods on the basis of the available evidence.

Methods

We collected information for this review from our own experience, from discussions with other experts in this field, and from Medline searches by using the search terms "assessment," "technical skills," "psychomotor skills," "competence," "surgery," "simulations," "dexterity," and "virtual reality." We cross referenced some of the information from other articles and proceedings and abstracts of papers presented at conferences.

Summary points

The assessment of technical skills is currently subjective and unreliable

Objective feedback of technical skills is crucial to the structured learning of surgical skills

Methods of assessment such as examinations, log books, and non-criteria based direct observation of procedures lack validity and reliability

Validated methods such as checklists, global rating scales, and dexterity analysis systems are suitable for the objective formative feedback of technical skills during training

Virtual reality systems have the potential to be used for assessment in the future

Further research is needed before these methods can be used for summative assessment and revalidation of surgeons

The surgical community could follow the example of other high reliability organisations such as aviation and design training programmes, where continuous assessment is a part of training

Current methods of assessment in surgery and their limitations

Any assessment method should be feasible, valid, and reliable (box 1).² Currently five methods are available for assessing technical skills that are valid and reliable to varying degrees (table 1).² It is evident that some of the methods of assessment currently in use have poor validity and reliability. Examinations such as the fellowship and membership of the Royal College of Surgeons (FRCS and MRCS), conducted jointly by the royal colleges, focus mainly on the knowledge and clinical abilities of the trainee and do not assess a trainee's technical ability. One study showed that no relation existed between the American Board of Surgery in Training Exam (ABSITE) score and technical skill.⁸ All

Table 1 Assessment of technical skills—validity and reliability

Method of assessment	Reliability	Validity
Procedure lists with logs	Not applicable	Poor
Direct observation	Poor	Modest
Direct observation with criteria	High	High
Animal models with criteria	High	Proportional to realism
Videotapes	High	Proportional to realism

trainees in the United Kingdom are required to maintain a log of the procedures performed by them, which is submitted at the time of examinations, at job interviews, and during annual assessments. However, it has been found that log books are indicative merely of procedural performance and not a reflection of operative ability and therefore lack content validity.^{1 2} Time taken for a procedure does not assess the quality of performance and is an unreliable measure when used during real procedures, owing to the influence of various other factors.

The assessment of technical skills by observation, as currently occurs in the operating room, is subjective. As the assessment is global and not based on specific criteria it is unreliable. As it is influenced by the subjectivity of the observer it would possess poor test-retest reliability and also be affected by poor interobserver reliability as even experienced senior surgeons have a high degree of disagreement while rating the skills of a trainee.²

Morbidity and mortality data, often used as surrogate markers of operative performance, are influenced by patients' characteristics, and it is believed that they do not truly reflect surgical competence.¹⁰

Objective methods of assessing technical skills

Checklists and global scores

The availability of set criteria against which technical skills can be assessed makes the assessment process more objective, valid, and reliable (box 2). It has been said that checklists turn examiners into observers, rather than interpreters, of behaviour, thereby removing the subjectivity of the evaluation process.¹¹ The wide acceptance of the objective structured clinical examination (OSCE) led a group in Toronto to develop a similar concept for the assessment of technical skills.⁹ The objective structured assessment of technical skills (OSATS) consists of six stations where residents and trainees perform procedures on live animal or bench models in fixed time periods.¹² Performance during the performance of tasks is assessed by using checklists specific to the operation or task (table 2) and a global rating scale (table 3). The global scale consists of seven generic components of operative skill that are marked on a 5 point Likert scale, with the middle and the extreme points anchored by explicit descriptors¹² to help in the criterion referenced assessment of performance. By using both formats of assessment Regehr et al have shown that checklists do not add any additional value to the assessment process and that their reliability is lower than that for the global rating scale.¹¹

Both live animal operating and bench models have been used for the OSATS assessment. No differences became obvious in the performance of trainees in both

Box 1: Principles of assessment

Validity

Construct validity is the extent to which a test measures the trait that it purports to measure. One inference of construct validity is the extent to which a test discriminates between various levels of expertise

Content validity is the extent to which the domain that is being measured is measured by the assessment tool—for example, while trying to assess technical skills we may actually be testing knowledge

Concurrent validity is the extent to which the results of the assessment tool correlate with the gold standard for that domain

Face validity is the extent to which the examination resembles real life situations

Predictive validity is the ability of the examination to predict future performance

Reliability

Reliability is a measure of a test to generate similar results when applied at two different points.² When assessments are performed by more than one observer another type of reliability test is applicable that is referred to as inter-rater reliability, which measures the extent of agreement between two or more observers⁹

formats of the examination.⁹ The interstation reliability for the six stations was also found to be high.¹²

The only drawback to the performance of the OSATS assessment are the resources and time involved in getting several staff surgeons to observe the performance of trainees. Retrospective video watching of the performance may be a way forward. It does not entail the presence of multiple faculty raters and also adds to the element of objectivity by making the assessment blinded. By using this method of assessment Datta et al showed the construct validity of the global rating scale with an inter-rater reliability of 0.81.¹³

The use of both checklists and global rating scales entail the presence of multiple faculty raters or extensive video watching. Systems therefore need to be developed that can produce an assessment of technical skills in real time, with little need for several observers. Dexterity analysis systems have the potential to address this issue.

Dexterity analysis systems

Imperial College surgical assessment device

This is a commercially available electromagnetic tracking system (Isotrak II, Polhemus, United States), which

Box 2: Methods of assessment in surgery

Current

- Examinations
- Operative log books
- Time taken for a procedure
- Direct observation and assessment by trainers
- Morbidity and mortality data

Recent developments

- Checklists
- Global rating scales, such as OSATS (objective structured assessment of technical skills)
- Dexterity analysis systems, such as ICSAD (Imperial College surgical assessment device), ADEPT (advanced Dundee endoscopic psychomotor trainer)
- Virtual reality simulators
- Analysis of the final product on bench models
- Error scoring systems

Table 2 Checklists

Small bowel anastomosis:
Interrupted end to end single layer anastomosis. Score one point for each correctly performed action

Tape No: _____ Item No: _____
Assessor: _____ (initials only)

Procedural step	Correctly performed	Incorrectly performed
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		

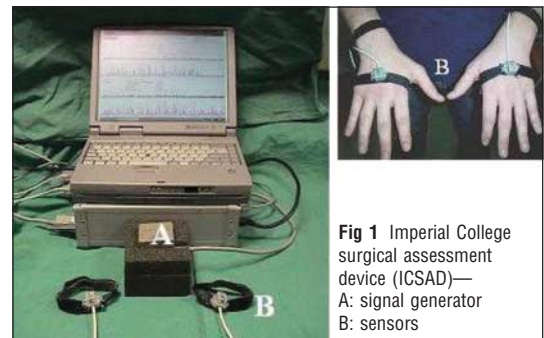


Fig 1 Imperial College surgical assessment device (ICSAD)—
A: signal generator
B: sensors



Fig 2 Typical trace from Imperial College surgical assessment device

consists of an electromagnetic field generator and two sensors that are attached to the dorsum of the surgeon's hands at standardised positions (fig 1). Bespoke software is used for converting the positional data generated by the sensors to dexterity measures such as the number and speed of hand movements, the distance travelled by the hands and the time taken for the task. Recently, our group has developed new software that allows us to stream video files along with the dexterity data in order to zoom into certain key steps of a procedure (fig 2).

Studies have shown the construct validity of the Imperial College surgical assessment device with respect to a range of laparoscopic and open surgical tasks.¹⁴ The correlation between dexterity and previous

laparoscopic experience on a simple task in a box trainer¹⁵ and on more complex tasks such as laparoscopic cholecystectomy on a porcine model is strong.¹⁶ Experienced and skilled laparoscopic surgeons are more economical in terms of the number of movements and more accurate in terms of target localisation and therefore use much shorter paths.

Other motion analysis systems

Motion tracking can be based on electromagnetic, mechanical, or optical systems. The advanced Dundee endoscopic psychomotor trainer (ADEPT) was originally designed as a tool for the selection of trainees for endoscopic surgery, based on the ability of psychomotor tests to predict innate ability to perform relevant tasks. Studies have shown the validity and reliability of the

Table 3 Global rating scale

Variable	Rating				
	1	2	3	4	5
Respect for tissue	Often used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately, with minimal damage
Time and motion	Many unnecessary moves		Efficient time and motion, but some unnecessary moves		Economy of movement and maximum efficiency
Instrument handling	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments, although occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
Knowledge of instruments	Frequently asked for the wrong instrument or used an inappropriate instrument		Knew the names of most instruments and used appropriate instrument for the task		Obviously familiar with the instruments required and their names
Use of assistants	Consistently placed assistants poorly or failed to use assistants		Good use of assistants most of the time		Strategically used assistant to the best advantage at all times
Flow of operation and forward planning	Frequently stopped operating or needed to discuss next move		Demonstrated ability for forward planning with steady progression of operative procedure		Obviously planned course of operation with effortless flow from one move to the next
Knowledge of specific procedure	Deficient knowledge. Needed specific instruction at most operative steps		Knew all important aspects of the operation		Demonstrated familiarity with all aspects of the operation

trainer.¹⁷ Optical motion tracking systems consist of infrared cameras surrounded by infrared light emitting diodes. The infrared light is reflected off sensors that are placed on the limb of a surgeon.¹⁸ Software is used to extrapolate the positional data of the markers to data on movement analysis. The disadvantages of such optical systems are that they suffer from disturbances to the line of vision. If the camera and signal generators become obscured from the markers the resulting loss in link leads to lost data. Signal overlap also prevents the use of markers on both limbs, making these systems restrictive.

Virtual reality

Virtual reality is defined as a collection of technologies that allow people to interact efficiently with three dimensional computerised databases in real time by using their natural senses and skills.¹⁹ Surgical virtual reality systems allow interaction to occur through an interface, such as a laparoscopic frame with modified laparoscopic instruments (fig 3).

The minimally invasive surgical trainer-virtual reality (MIST-VR) system was one of the first virtual reality laparoscopic simulators developed as a task trainer (fig 3). The system was developed as the result of collaboration between surgeons and psychologists who performed a task analysis of laparoscopic cholecystectomy. This resulted in a toolkit of skills needed to perform the procedure successfully.¹⁹ These were then replicated in the virtual domain by producing three dimensional images of shapes, which users can manipulate.

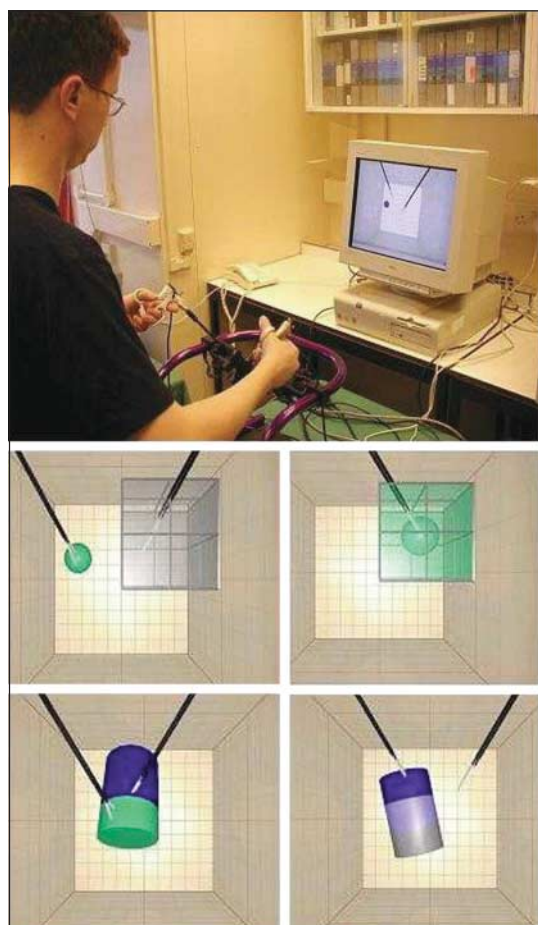


Fig 3 Minimally invasive surgical trainer-virtual reality (MIST-VR)

As virtual reality simulators are computer based systems they generate output data, or what is commonly referred to as metrics. In an international workshop a group of experts reviewed all methods of assessment and suggested parameters that should constitute output metrics for the assessment of technical skills.²⁰ Included were variables such as economy of movement, length of path, and instrument errors. The MIST-VR system has been validated extensively for the assessment of basic laparoscopic skills.²¹ As these systems are currently low fidelity task trainers they are effective only in the assessment of basic skills. In the future, however, there is a possibility that higher fidelity virtual reality systems may be used as procedural trainers and for the assessment of procedural skill. One of the main advantages of virtual reality systems, in comparison to dexterity analysis systems, is that they provide real time feedback about skill based errors.

Analysis of the final product

As outcomes after surgery are often difficult to ascribe solely to surgical technique, and as adverse outcomes from poor technique may not be apparent for many years—for example, recurrence after cancer resections—some researchers have suggested the idea of using outcome measures on bench models. Szalay et al assessed the quality of the final product after performance of six different bench model tasks.²² The investigators found that the method possessed construct validity. They also found a correlation between OSATS and the final product assessment, which implies that analysis of the final product may overcome some of the problems involved with live ratings.

Datta et al assessed the leak rates and cross sectional area of the lumen after performance of a vascular anastomosis on a bench model and found a significant correlation between these outcome measures and surgical dexterity.²³ Hanna et al studied the quality of knots performed laparoscopically by using a tensiometer and derived a quality score for knots as an index of knot reliability.²⁴ The main advantage of these outcome measures is that they address the limitations of live and video based assessments and can be combined with dexterity data to derive proficiency scores, making assessment much more objective.

Discussion

The previous section has highlighted the various methods that have been developed over the past decade for the objective assessment of technical skills in surgery. Educational bodies such as the Joint Committee for Higher Surgical Training (JCHST) in general surgery in the United Kingdom have appreciated the need to increase the emphasis laid on the assessment of technical skills during training. Hence the committee has recommended the use of an assessment of operative competence, with consultant surgeons assessing their trainees for procedures performed during a fixed training period. The assessment uses five overall global ratings (box 3) to rate the trainee's ability to carry out procedures.

However, a crucial issue that needs to be addressed regarding the assessment environment is whether assessments should be carried out on simulations with adequate face validity or performed during real proce-

Box 3: Assessment of operative competence

- U: Unknown (not assessed) during the training period
- A: Competent to perform the procedure unsupervised (can deal with complications)
- B: Does not usually require supervision but may need help occasionally
- C: Able to perform the procedure under supervision
- D: Unable to perform the entire procedure under supervision

dures. One drawback of the latter approach is that it is impossible to ensure standardisation because all patients are different, and trainers can be more preoccupied with patients' safety and timely completion of the procedure than with concentrating on the details of operative skill and technique. Further research is required regarding the feasibility and reliability of assessment during real procedures.

Animal laboratories are a common and popular method of learning skills in North America and Europe, but their use is banned in the United Kingdom. The use of synthetic models or simulations for the acquisition of basic technical skills is becoming increasingly acknowledged. Research groups such as ours have shown the validity of synthetic models for both training and assessment. Such simulations, while being crucial for learning, can also be used to assess skills simultaneously. In fact, assessment and training are synergistic. Without objective, valid, and reliable assessment training programmes cannot ensure the learning of skill, tackle deficiencies in training, and implement remedial measures. Figure 4 shows a recommended format for a cycle of training and assessment that could ensure that progression from one level to the higher one is based on robust criteria. This is based on the model of continuous training and assessment of pilots. Table 4 shows a suggested panel of tasks on bench models and virtual reality simulators according to the level of training. Such a paradigm shift in surgical learning will also ensure that trainees get objective feedback throughout their training programmes.

Other applications of technical skills assessment

In addition to the assessment of skills during training, measures described in this article can be used to show the efficacy of training courses in teaching the participants psychomotor skills. This would also give participants in the course an opportunity to gain an insight into the skills that they have learnt and allow training

centres to strive to improve the quality of teaching, ensure standardisation, and change course formats according to the performance of participants. The benefit of being able to measure surgical skill also presents the surgical community with the opportunity to show objectively the effect of training interventions, such as the use of virtual reality simulators, and to study the effect on technical performance of adverse environmental conditions, such as sleep deprivation, and distractions, such as noise in the operating theatre.

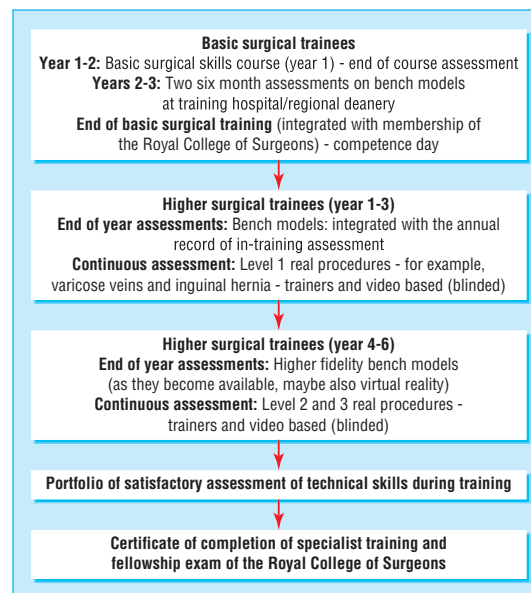


Fig 4 Recommended format for a cycle of training and assessment

The objective assessment of skills inevitably raises the issue of summative assessment and revalidation. Our group has previously alluded to the creation of a competence day for junior surgical trainees as an examination format by using OSATS and ICSAD (table 4). However, certain challenges (box 4) will have to be addressed before this becomes feasible for senior surgical trainees and consultant surgeons.²⁵

Conclusion

We have attempted to highlight the considerable progress made in the past decade in the objective assessment of technical skills. The surgical community can choose from a wide range of methods of assessment initially during training to make feedback

Table 4 Suggested panel of tasks

Level of training	Tasks and methods of assessment
Basic surgical trainees (junior residents)	Knot tying (ICSAD)
	Suturing (simple, mattress, and precision)—ICSAD
	Excision of skin lesions—for example, sebaceous cyst (OSATS)
	Small bowel enterotomy (OSATS)
	Basic laparoscopic skills (virtual reality simulators)
Higher surgical trainees, years 1-3 (middle level residents)	Small bowel anastomosis (ICSAD, OSATS)
	Vein patch insertion (OSATS)
	Saphenous vein dissection and ligation (OSATS)
	Basic laparoscopic skills (virtual reality simulators)
	Basic endoscopy skills (virtual reality simulators)

ICSAD=Imperial College surgical assessment tool.
OSATS=objective structured assessment of technical skills.

Box 4: Research currently in progress

- Feasibility, validity, and reliability of objective assessment during real procedures using ICSAD and video based assessment
- Studies to explore a link between technical skills and outcome for patients
- Studies addressing the predictive validity of objective assessment methods
- Correlation between performance on simulations and real procedures
- Longitudinal studies using a large cohort of trainees followed over a period of time to evaluate the link between training and assessment
- The establishment of databases for different tasks and procedures to establish performance criteria in order to make progress during training criteria based
- Use objective measures of surgical skills to demonstrate transfer of skills from virtual reality to real procedures
- Assessment of surgical competence in realistic environments such as a simulated operating theatre for assessment of communication, decision making, and leadership, and for training and assessment of crisis management

more objective, to base progression on criteria, and to help poorly performing trainees take remedial action. There are, however, still some issues that need to be addressed (box 4).

In addition to being of crucial importance for training, technical skills assessment is also driven by the need for the surgical community to ensure surgical care of the highest quality and reduce any potential errors resulting from poor technical performance. Owing to our inability to measure surgical skills objectively, so far it has been difficult to show a link between technical performance and outcome for patients. Future research should try to explore the link between technical skills assessed objectively and postoperative measures such as complication and recurrence rates and postoperative pain.

It must, however, be emphasised that technical skills are only a part of surgeon's competence, and the assessment of technical skills needs to be integrated with

cognitive and behavioural characteristics such as team skills and decision making in order to develop methods that assess surgical competence comprehensively.

Contributors: All authors participated in the preparation of the manuscript and in the revision process. AD is the guarantor.

Funding: BUPA Foundation.

Competing interests: None declared.

- 1 Cuschieri A, Francis N, Crosby J, Hanna GB. What do master surgeons think of surgical competence and revalidation? *Am J Surg* 2001;182:110-6.
- 2 Reznick RK. Teaching and testing technical skills. *Am J Surg* 1993;165:358-61.
- 3 Kopta JA. An approach to the evaluation of operative skills. *Surgery* 1971;70:297-303.
- 4 Bridges M, Diamond DL. The financial impact of teaching surgical residents in the operating room. *Am J Surg* 1999;177:28-32.
- 5 Ross DG, Harris CA, Jones DJ. A comparison of operative experience for basic surgical trainees in 1992 and 2000. *Br J Surg* 2002;89(suppl 1):60.
- 6 Skidmore FD. Junior surgeons are becoming deskilled as result of Calman proposals. *BMJ* 1997;314:1281.
- 7 Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. *BMJ* 1999;318:887-8.
- 8 Scott DJ, Valentine RJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, et al. Evaluating surgical competency with the American board of surgery in-training examination, skill testing, and intraoperative assessment. *Surgery* 2000;128:613-22.
- 9 Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;84:273-8.
- 10 Bridgewater B, Grayson AD, Jackson M, Brooks N, Grotte GJ, Keenan DJ, et al. Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data. *BMJ* 2003;327:13-7.
- 11 Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998;73:993-7.
- 12 Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg* 1997;173:226-30.
- 13 Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. *Am J Surg* 2002;184:70-3.
- 14 Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg* 2001;193:479-85.
- 15 Taffinder N, Smith S, Mair J, Russell R, Darzi A. Can a computer measure surgical precision? Reliability, validity and feasibility of the ICSAD. *Surg Endosc* 1999;13(suppl 1):81.
- 16 Smith SG, Torkington J, Brown TJ, Taffinder NJ, Darzi A. Motion analysis. *Surg Endosc* 2002;16:640-5.
- 17 Francis NK, Hanna GB, Cuschieri A. The performance of master surgeons on the advanced Dundee endoscopic psychomotor tester: contrast validity study. *Arch Surg* 2002;137:841-4.
- 18 Emam TA, Hanna GB, Kimber C, Cuschieri A. Differences between experts and trainees in the motion pattern of the dominant upper limb during intracorporeal endoscopic knotting. *Dig Surg* 2000;17:120-3.
- 19 McCloy R, Stone R. Science, medicine, and the future. Virtual reality in surgery. *BMJ* 2001;323:912-5.
- 20 Satava RM, Cuschieri A, Hamdorf J. Metrics for objective assessment. *Surg Endosc* 2003;17:220-6.
- 21 Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A. 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.
- 22 Szalay D, MacRae H, Regehr G, Reznick R. Using operative outcome to assess technical skill. *Am J Surg* 2000;180:234-7.
- 23 Datta V, Mandalia M, Mackay S, Chang A, Cheshire N, Darzi A. Relationship between skill and outcome in the laboratory-based model. *Surgery* 2002;131:318-23.
- 24 Hanna GB, Frank TG, Cuschieri A. Objective assessment of endoscopic knot quality. *Am J Surg* 1997;174:410-3.
- 25 Darzi A, Datta V, Mackay S. The challenge of objective assessment of technical skill. *Am J Surg* 2001;181:484-6.

(Accepted 4 September 2003)

Additional educational resources**Reviews**

Darzi A, Datta V, Mackay S. The challenge of objective assessment of surgical skill. *Am J Surg* 2001;181:484-6.

Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Assessment of technical surgical skill. *Eur J Surg* 2002;168:139-44.

Moorthy K, Munz Y, Dosis A, Bello F, Darzi A. Motion analysis in the training and assessment of laparoscopic surgery. *Min Invas Ther Allied Technol* 2003;12:37-42.

Websites

www.jchst.org—Joint Committee of Higher Surgical Training, for further details of the assessment of operative competence for higher surgical trainees in the United Kingdom

www.surgicaleducation.com—Association of Surgical Education, dedicated to promoting the art and science of education in surgery

Endpiece**Hard people**

Writers, like teeth, are divided into incisors and grinders.

Walter Bagehot (1826-77), English economist, social scientist, and journalist in *The first Edinburgh reviewers*, 1879

Fred Charatan, retired geriatric physician, Florida