January 2009

Assessment, surgeon, and society

John Norcini

Jamsheer J Talati
Aga Khan University, jamsheer.talati@aku.edu

International Journal of Surgery 7 (2009) 313–317

Contents lists available at ScienceDirect

International Journal of Surgery

journal homepage: www.theijs.com

Review

Assessment, surgeon, and society

John Norcini [a,*], Jamsheer Talati [b]

[a] Foundation for Advancement of International Medical Education and Research, United States
[b] The Aga Khan University, Pakistan

ARTICLE INFO

Article history:
Received 8 June 2009
Accepted 23 June 2009
Available online 30 June 2009

Keywords:
Assessment
Education
Surgeon

ABSTRACT

An increasing public demand to monitor and assure the quality of care provided by physicians and surgeons has been accompanied by a deepening appreciation within the profession of the demands of self-regulation and the need for accountability. To respond to these developments, the public and the profession have turned increasingly to assessment, both to establish initial competence and to ensure that it is maintained throughout a career. Fortunately, this comes at a time when there have been significant advances in the breadth and quality of the assessment tools available. This article provides an overview of the drivers of change in assessment which includes the educational outcomes movement, the development of technology, and advances in assessment. It then outlines the factors that are important in selecting assessment devices as well as a system for classifying the methods that are available. Finally, the drivers of change have spawned a number of trends in the assessment of competence as a surgeon. Three of them are of particular note, simulation, workplace-based assessment, and the assessment of new competences, and each is reviewed with a focus on its potential.

> He is a good surgeon who possesses courage and presence of mind, a hand free from perspiration, tremorless grip of sharp and good instruments, and who carries his operation to success and to the advantage of his patient, who has entrusted his life to the surgeon. The surgeon should respect the absolute surrender and treat his patient as his own son.
> Sushruta
> RK Choudhary, *staff grade surgeon* BMJ 2004;329:314 (7 August).

The monitoring and punishment of negligent or errant physicians have been in existence for centuries; and have at times been dictated by the ruling authority, often in an inequitable manner, as in the code of Hammurabi.[1] More recently, there have been increasing demands to monitor and assure the quality of care provided by physicians and surgeons. Many of these demands have been driven by high profile cases such as the one at the British Royal Infirmary in the UK where three medical practitioners were found guilty of professional misconduct relating to 29 deaths in 53 pediatric cardiac operations.[2] Accompanied by other high profile cases such as that of Swango in the US and Shipman in the UK, patients and the public have called for additional scrutiny of doctors.[3,4]

At the same time that the public has been clamoring for change, there has been a deepening appreciation within medicine of the demands of self-regulation and the need for professionalism and accountability.[5] With the assumption of responsibility, it is no longer acceptable to rely only on the unstructured judgments of colleagues to assure the public that surgeons can provide safe and effective care. This does not fulfill the obligations of a self-regulating profession, it does not serve the needs of patients, and it invites closer governmental regulation of surgical training and practice.

To respond to these developments, the public and the profession have turned increasingly to assessment both to establish initial competence and to ensure that it is maintained throughout a career. Fortunately, at the same time there have been significant advances in the breadth and quality of the assessment tools available.[6,7]

This paper is one of the occasional articles in a series on assessment and its goal is to provide an overview of 1) the drivers of change in assessment, 2) factors in selecting assessment devices, 3) a system for classifying methods of assessment along with some examples, and 4) future trends.

1. Drivers of change in assessment

Historically, the assessment of physicians and surgeons was based on a limited set of written and oral examination methods. In turn, these methods severely limited the competences that could

* Correspondence to: John Norcini, Foundation for Advancement of International Medical Education and Research, 3624 Market Street, 4th Floor, Philadelphia, PA 19104, United States. Tel.: +1 215 823 2170.
E-mail address: jnorcini@faimer.org (J. Norcini).

1743-9191/$ – see front matter
doi:10.1016/j.ijsu.2009.06.011

be assessed and the quality of the decisions that could be made. More recently, three important trends have converged to increase the number and sophistication of the tools available to establish accountability.[7] First, the traditional emphasis on the process of education has gradually been replaced with a focus on the competences of surgeons at the end of training (i.e., outcomes) and these competences have guided the development of new methodologies. Second, increased technological sophistication has impacted both the development of new methods of assessment as well as the delivery of traditional methods. Third, there have been advances in psychometrics, the basic science of assessment, which have increased both the quality and efficiency of measurement.

1.1. Educational outcomes movement

Medical education has followed the trends in general education away from an emphasis on how students are taught to what is expected from them by the end of an educational experience. For example, medical education in the US has coalesced around six competences: 1) medical knowledge, 2) interpersonal skills, 3) patient care, 4) professionalism, 5) systems-based practice, and 6) practice-based learning and improvement.[8] These same competences apply equally to the medical and surgical specialties. Similarly, the General Medical Council in the UK has settled on 1) good clinical care, 2) maintaining good medical practice, 3) teaching and training, appraising and assessing, 4) relationships with patients, 5) working with colleagues, 6) probity, and 7) health as the competences of the good doctor.[9] Having been specified, these outcomes drive the curriculum at all levels of training, the nature of the teachers, and even the sites of training.

This shift in focus also changes the nature of assessment.[10] Historically, testing has been aimed at determining whether students learned what was taught. As the emphasis shifts away from the process of education, assessment assumes the more central role of determining whether students have mastered the specified outcomes. In fact, this central role for assessment has led some to suggest that the duration of the curriculum could be variable as long as students achieved the appropriate outcomes.

1.2. Development of technology

The past fifty years have seen dramatic growth in technology and these developments have been reflected in assessment. At first, the technology was used for test administration tasks such as scoring and reporting of results. With growing sophistication, however, it became useful in simulating patient conditions and procedures.[11] This made it possible to train physicians and surgeons in an environment where errors caused no harm to patients and for problems which were infrequently seen in practice.

Naturally, the availability of simulation for education has led to its use in assessment as well. It has permitted the development of more valid assessments, supported direct comparisons among test takers (since they can all be exposed to identical test material), and, as with the educational use of this technology, allowed the assessment of complex skills in a safe environment.

1.3. Psychometric advances

Psychometrics is the basic science of assessment. The rapid changes in technology have been accompanied by equally rapid changes in psychometrics. Classical test theory, in use from the early 1900s gave way to item response theory on the one hand, and Generalizability theory on the other.

The Item Response theory family of models makes strong assumptions about the assessments to which they are applied.[12] However, if they are met it is possible to produce equivalent scores even when different sets of items are administered or when they are scored by different groups of examiners. Moreover, these models support the administration of assessments that adapt to the ability of students (becoming harder for students of higher ability and easier for students of lower ability) which makes for shorter assessments.

Generalizability theory permits the user to more fully understand the sources of error in an assessment and thereby provides guidance for minimizing them given the available resources.[13] For instance, in an oral examination, the examiner(s) are one source of error and the patient(s) are a second. Application of Generalizability theory can help to determine the size of the overall errors of measurement and whether increasing the number of examiners or patients will have the greatest impact in reducing them.

2. Factors in selecting methods of assessment

In many postgraduate training programs, different assessment methods are treated as though they stand alone. Although it is not often explicitly acknowledged, important decisions are usually based on information gained from different components within a system of assessment.[7] Attention to the design of this system, its purposes, and the interactions among its components is essential to achieving good results. The assessment system that develops in any surgical program will result from a collection of assessment methods and devices that serve the objectives one wishes to achieve.

Given that there is a system design, methods of assessment for a particular purpose have traditionally been chosen based on their validity and reliability. Validation is the process of gathering data that supports the inferences made on the basis of the assessment – is it fit for purpose? As research on the validity of the traditional assessments grew, it became clear that new methods were needed. For instance, a written examination provides an excellent means of assessing the indications and contraindications for a procedure but not of the skill required to perform it. Consequently, improvement of validity became a motivating force for the development of other methods, such as simulation.

Reliability refers to the reproducibility of scores from an assessment. If a test is repeated we would ideally expect students to get exactly the same score. The degree to which scores are the same is referred to as reliability and there are several indices for evaluating it. On the one hand, since it is relatively easy to calculate, reliability is often the primary measure of test quality. Validity, however, is more important. On the other hand, a test that is not reliable cannot be valid either.

Recently, van der Vleuten and Schuwirth have added three other factors that should be taken into account when determining the utility of a method for a particular purpose: feasibility, acceptability, and educational effect.[14] Feasibility refers to whether the method is doable given available resources of faculty and time, and whether it can be deployed in an efficient and reliable fashion. For example, the traditional clinical viva might be a preferred method of assessment for a particular purpose, but it may be impractical to include enough different examiners and patients to produce reliable results.

Acceptability refers to whether the faculty and students believe that the method will produce credible results. If not, faculty will not be motivated to use it and students, especially those adversely affected, will not accept the results.

Finally, students will work hard to pass the assessments that are in place – they learn to the test. It is therefore important that the assessment methods be selected so as to motivate students in ways that are consistent with learning objectives. For example, if the goal

is for students to learn how to take a good history from a patient then assessment based on standardized patients will be more effective than assessment based on an essay examination.

3. Methods of assessment

A review of the literature will reveal a vast variety of methods of assessment; and the numbers are growing continually. To put some order to this, George Miller, an eminent physician educator, proposed a classification scheme based on what was required for trainees.[15]

At the lowest level is 'knows'. As the name implies, these are the methods that determine how much an examinee knows. Multiple choice questions (MCQs) are the most popular example for determining how much knowledge a surgeon possesses but essays, oral examinations, and simulations may be used for this as well but none are likely to be as efficient as MCQs. This is particularly true because of their ability to scan large fields of knowledge in limited time, something that has become increasingly possible by electronic testing.

The next level is 'knows how'. These are methods which test the ability to assemble the facts in service of a clinically relevant purpose. Knowing how to conduct a particular procedure or developing a differential diagnosis given a specific set of presenting complaints are examples. Again, MCQs can be used to test this type of competency as can essays, oral examinations, and more recently, the script concordance test and a variety of different simulations. Likely however, well-constructed MCQs will be more efficient.

The third level is 'shows how' or demonstrates a skill. For example, having students take a history from a patient or asking them to perform an end-to-end anastomosis on a model would be examples of 'showing how'. Methods such as standardized patients, oral examinations, and various low to high fidelity simulations are exemplars of this level of the pyramid. Written methods, like MCQs, are only rarely useful in 'showing how' as they can only test the cognitive components necessary for the performance of the task, which may be performed incorrectly or in a risky manner despite adequate knowledge.

The top level of the pyramid is 'does' and it refers to routine performance, not what happens when the surgeon knows he/she is being assessed. As a consequence, most of the common methods like written exams, oral exams, and simulations are ruled out. Instead for example, assessments based on patient outcomes (e.g., severity-of-illness adjusted mortality and morbidity rates) are appropriate at this level of the pyramid, as are methods such as peer and patient surveys.

4. The future

The drivers of change mentioned above have spawned a number of trends in the assessment of competence as a surgeon. Three of them are of particular note: simulation, workplace-based assessment, and the assessment of new competences.

4.1. Simulation

Simulation is increasingly being used in the assessment of physicians and surgeons.[11] For example, licensure in the US now requires successful performance on examinations that include both standardized patients (actors trained to play the role of patients) and computer-based simulations. These newer technologies are well developed and they provide an assessment of essential components of competence.

Within surgery, two trends in simulation are particularly worthy of note. First, with the increasing sophistication of computers has come the ability to recreate with high fidelity various patient conditions and procedures. Simulation has become a basic tool in the education of doctors but it also provides advantages for assessment.[16] Trainees can be assessed on serious and/or infrequent medical–surgical problems in safety. Moreover, it is possible to directly compare scores since each trainee can be presented identical 'patients' and the high fidelity of the challenges enhances the validity of the scores. As the number and quality of methods for simulating with high fidelity grow, use of these tools for assessment will increase.

Also worthy of note are hybrid models that integrate standardized patients with manikins or other simulators to test both communication skills and a variety of basic procedural skills in the context of a patient encounter.[16] For example, trainees might be expected to suture or insert a urinary catheter into a model while simultaneously interacting with the standardized patient. The standardized patient can be trained to demonstrate a range of challenging behaviors during the 'procedure' so that the candidate's response to these can be evaluated.

As the technology continues to improve, it is reasonable to believe that the use of simulation will grow. This is particularly true in surgery, where its use for safe and effective procedural skills training and assessment will be a significant advance.

4.2. Workplace-based assessment

The past several decades have seen considerable improvement in the assessment of cognitive and clinical skills. As much of the effective education of a surgical resident occurs in the workplace, there still remains a need for methods that assess their ability to work in teams; make correct judgments based on accurate observation, examination, and clinical reasoning; communicate difficult and divergent options to patients and their relatives; tackle challenging unexpected turns of events in day to day work; and recognize the 'tipping points'[17] that will demand wide overall change in the ways we provide care. While quite powerful, the OSCE is not ideally suited to assess these complexities of clinical practice.

The methods used should be feasible and suitable for use in clinical settings.[19] Moreover, as the level of training increases, the assessments need to pose trainees a broader and often more acute set of patient problems. It is also important that they both fit into and support the educational mission of the workplace setting.[18] Further, as the quality of the residents' training is dependent on the effectiveness of the surgeon-teachers in their roles as educators, assessors, as well as proficient professional (surgeon) service providers, the evaluation of these qualities in the surgeon-teacher forms an important component of the workplace-based assessment.

To address the lack of good options in the workplace context, a group of assessment methods have been proposed. They capitalize on two aspects of the setting.[18] First, trainees are in routine interactions with patients, peers, and other members of the healthcare team. These interactions form a significant body of clinical material that can serve as the stimulus for assessment. Second, clinician educators in these settings can act as examiners and provide feedback. Hence, many of the methods proposed are based on faculty observation of routine encounters. As many clinical settings are supervised by relatively few faculty members and the curriculum is often not specified in detail, some schools depute tutors, who assist with provision of an appropriate learning environment, and ensure that standards relevant to education are understood and met.[19]

The four methods used as part of the Foundation Programme in the United Kingdom serve as good examples of workplace-

based assessment structured around the observation of performance: mini-Clinical Evaluation Exercise (mini-CEX), Direct Observation of Procedural Skills (DOPS), Case-based Discussion (CbD), and Multi-Source Feedback (MSF).[20] In the mini-CEX and DOPS, the trainee is observed in a patient encounter performing specific tasks (history/physical exam for the mini-CEX and procedures for DOPS) and then assessed and given feedback. In CbD, the trainee and faculty member explore a patient record in which the trainee has made notes. At the end of the discussion, the trainee is assessed and given feedback. Finally, MSF is an anonymous collection of assessments, gathered by rating form, from peers and colleagues.

Workplace-based assessment methods are simply systematic ways of collecting information for assessment and providing feedback. They rely heavily on the judgment of clinician educators and, as such, will require some faculty development.[21]

4.3. New competences

In a rapidly changing world, each surgeon needs to recognize, manage and channel changes for the benefit of patients (and self). As an example, each surgeon should keep abreast of developments in surgical technique and technology; evaluate their usefulness and courageously abandon old fashioned techniques when the results of newer techniques are superior.[22]

The shift in emphasis from educational process to educational outcomes has led to a broadening of the definition of competence. This expanded definition means that methods of assessment need to be developed and refined to capture competence in these newer areas of which professionalism is an example.

In recent years there has been increased interest in professionalism.[2–6] From the perspective of patients, this trend is not only driven in part by the extreme cases of misconduct such as those cited above but also by the more common concerns about the insensitivity, impatience, and inattentiveness of some physicians and surgeons. From the doctors' perspective, it expresses the profession's desire to avoid egregious problems and to promote routine appropriate behavior.

The development of methods to assess professionalism is in its infancy.[6] Nonetheless, a variety of alternatives exist. These range from MCQs that assure a basic fund of knowledge, through methods that test whether a trainee knows how to solve dilemmas in professionalism, on to methods, such as faculty observation that require students to show how they deal with patients in this domain, to peer and patient assessments which capture routine behaviors.

Over the next decade, better ways to assess professionalism will emerge. Likewise, methods for use with some of the other newer competences like systems-based practice and practice-based learning will emerge.

5. Summary and conclusion

The importance of a comprehensive assessment system blueprinted to the curriculum and set to standards that ensure required competences cannot be overemphasized. Information derived from such systems improves postgraduate education, and assures society of the quality of its future surgeons. Individual assessments within the system must be reliable, valid, and probe thinking. Standards which need to be set carefully, could consider, but should not be based solely on peer performance.[23] Measurement should replace unstructured judgments, even in the field of operative skills.[24]

Assessment must be performed efficiently – the enthusiastic pursuit of comprehensiveness should not result in work overload for students or assessors; nor should complex methods be used to simply test acquisition of information. Each assessment must add value.

The greatest value of assessment lies in assessment of actions in the playing fields, where society demands practical demonstration of excellence, in everyday practice, over an extended period of time. As assessment supplies evidence for the state of the candidate at one point in time, and as the long term effect of education cannot be predicted,[25] systems must include methods for repeated assessment throughout life.

In addition to professional performance, many other competences deserve evaluation too. Amongst them prominently, are the ability to teach; the ability to adapt to and participate in the management of change; the ability to think broadly, adapt to the environment and resources, improvise or innovate, in order to develop less laborious or low cost, yet equally effective management protocols and techniques.[25] Important in low income countries, the development of these competences requires social awareness and interaction, and immersion in the culture of the city (as recommended in the Bologna Process[26]) and networking across disciplines. The assessment of the degree to which these underlying capacities have been absorbed to benefit the individual, adds yet another complex dimension to the world of assessment. They merit a separate consideration. Will immersion into society and the environmental context automatically generate such thinking, and make redundant the need for assessment? Unfortunately not. Assessment is required because a century of evidence shows that innovations are not occurring from the environs needing them most. Malaria was researched by a British Army surgeon; AIDS is studied not in Africa but in the USA and Europe.

Assessment is a crucial educational tool. It cannot rest alone in a silo. Assessment has to dance in rhythms that flow into the life line of teaching and learning. In response, the total educational experience has to be constantly modified to ensure that graduates have the potential to acquire the competences that will be required in the future. Many assessments such as the reflective log book, and record of personal growth of understanding of a problem and its solutions, are great learning experiences. The pas de deux between the two equally tempered thrusts –learning and assessment – will assume new dimensions as society demands perfection in performance at a time when educational strategies increasingly involve workplace-based learning.

As new methods evolve, each will itself require careful assessment of attainment of its purpose and the economy of effort – innovations without rigorous evaluation are of limited usefulness.[27] Each surgeon will have to ask, as did Rowley, "*how can the required complex mix of skills and attitudes be fairly judged?*" [28] and as did van der Vleuten,[14] 'How can it be done with a rational expenditure of energy?'

These are important questions. For, whether a surgeon is a University faculty or in consulting practice, the need to assess, form a composite picture of team members, and provide feedback in ways that enhance the individual and the quality of education, will be equally keenly felt.

In addition to knowledge base, the good surgeon of today must be tested to determine if s/he has i) the ***courage***, integrity, and ***presence of mind*** to respond to changes in the patient's illness or life in general; ii) a ***tremorless grip*** and economy of motion for expertly performing operations at speed, demonstrated by performance in the skill lab, and iii) her/his actions show that s/he thinks its his/her responsibility to listen and ***completely surrender self*** *in a professional manner* to justify the ***trust*** of patients and populations.

These are the same qualities which Sushruta suggested in antiquity. These must be important; they are the same qualities that

have been echoed by Guy de Chauliac,[29] who added that the surgeon 'should be ingenious and able to adapt himself', something that Engel[30] has drawn our attention to as *the* need for all practitioners for our evolving, changing century.

Conflict of interest
None declared.

Funding
None declared.

Ethical approval
None declared.

References

1. Code of Hammurabi, http://www.wsu.edu/~dee/MESO/CODE.HTM [last accessed 19.05.09].
2. Bristol case surgeon claimed to have been on learning curve. *Br Med J* 1999 December 4;**319**(7223):1456.
3. Stewart JB. *Blind eye: how the medical establishment let a doctor get away with murder*. New York City: Simon and Schuster; 1999.
4. Dyer O. Shipman murdered more than 200 patients, inquiry finds. *Br Med J* 2002 July 27;**325**(7357):181.
5. ABIM Foundation, ACP–ASIM Foundation, European Federation of Internal Medicine. Medical professionalism in the new millennium: a physician charter. *Ann Intern Med* 2002 February;**136**(3):243–6.
6. Stern DT, editor. *Measuring medical professionalism*. New York: Oxford University Press; 2006.
7. Holmboe ES, Hawkins RE, editors. *Practical guide to the evaluation of clinical competence*. United States: Elsevier Health Sciences; 2008.
8. Leach DC. A model for GME: shifting from process to outcomes. A progress report from the Accreditation Council for Graduate Medical Education. *Med Educ* 2004;**38**(1):12–4.
9. General Medical Council. *Good medical practice*. London: General Medical Council; 2001.
10. Harden RM, Crosby JR, Davis M. An introduction to outcome-based education. *Med Teach* 1999;**21**(1):7–14.
11. Tekian A, McGuire CH, McGaghie WC, editors. Innovative simulations for assessing professional competence. Chicago, IL: University of Illinois; 1999.
12. Hambleton RK, Swaminathan H. *Item response theory: principles and applications*. Dordrecht: Kluwer; 1985.
13. Brennan RL. *Generalizability theory*. New York: Springer-Verlag; 2001.
14. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005;**39**(3):309–17.
15. Miller G. The assessment of clinical skills/competence/performance. *Acad Med* 1990;**65**(suppl):S63–7.
16. Kneebone R. Simulation in surgical training: educational issues and practical implications. *Med Educ* 2003;**37**:267–77.
17. Malcolm Gladwell. The tipping point: how little things can make a big difference. Little Brown, http://en.wikipedia.org/wiki/The_Tipping_Point re; 2000.
18. Norcini JJ. Workplace-based assessment in clinical training. In: Swanwick T, editor. *Understanding medical education series. Association for the study of medical education*; 2007.
19. Spencer L. Who does what in Postgraduate Medical Education, http://www.yorksandhumberdeanery.nhs.uk/surgery/documents/UrologyPostgradMed.doc [last accessed 19.05.09].
20. Davies H, Archer J, Southgate L, Norcini J. Initial evaluation of the first year of the Foundation Assessment Programme. *Med Educ* 2009;**43**:74–81.
21. 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**(11):874–81.
22. Satava RM, Jones SB. Preparing surgeons for the 21st century. Implications of advanced technologies. *Surg Clin North Am* 2000;**80**(4):1353–65.
23. Postgraduate Medical Education and Training Board, UK (PMETB). Generic Standards for training, Standards for curriculum and assessment systems, http://www.pmetb.org.uk/index.php?id=scas [last accessed 15.05.09].
24. Jaffer B, Bednarz B, Challacombe, Sriprasad S. The assessment of surgical competency in the UK. *Int J Surg* 2009;**7**:12–5.
25. deCossart L, Fish D. *Cultivating a thinking surgeon: new perspectives on clinical teaching learning and assessment*. Shrewsbury: TFM Publishing Limited; 2005.
26. The Lourtie report. From Bologna to Prague: furthering the bologna process, Report to the ministers of education, http://www.bologna-bergen2005.no.
27. Barrows HS. Innovations without appropriate assessment are of limited usefulness. *Teach Learn Med* 2008;**20**:287.
28. Rowley DI. The surgeon's job: how should we assess the trainee? *J R Soc Med* 2004;**97**:363–5.
29. Guy de Chauliac. Antiques digest, http://www.oldandsold.com/articles11/medicine-16.shtml.
30. Engel CE. Health professions education for adapting to change and for participating in managing change. *Educ Health Change Learn Pract* 2000;**13**(1):37–44.