

## Assessment Methods in Medical Education

Medical education, the art and science behind medical learning and teaching, has progressed remarkably. Teaching and learning have become more scientific and rigorous, curricula are based on sound pedagogical principles, and Problem Based and other forms of active and self directed learning have become the mainstream. Teachers have progressed from the role of problem-identifier to that of the solution-provider.

During the last three decades medical schools have been faced with a variety of challenges from society, patients, doctors and students. They have responded in several ways including the development of new curricula, the introduction of new learning situations, the introduction of the new methods of assessment and a realization of the importance of staff development. Many effective and interesting innovations have been forthcoming.

The effective and efficient delivery of healthcare requires not only knowledge and technical skills but also analytical and communication skills, interdisciplinary care, counseling, evidence- and system-based care. This warrants our assessment systems to be comprehensive, sound and robust enough to assess the requisite attributes along with testing for essential knowledge and skills.

Assessment is entering every phase of professional development. Assessment and evaluation are crucial steps in educational process. Before making a choice of assessment method, some important questions must be asked: what should be assessed?, why assess? For an assessment instrument one must also ask: is it valid? Is it reliable?, is it feasible? What is assessed and which methods are used will play a significant part in what is learnt. A wide range of assessment methods currently available include essay questions, patient management problems, modified essay questions (MEQs) checklists, OSCE, student projects, Constructed Response Questions (CRQs), MCQs, Critical reading papers, rating scales, extended matching items, tutor reports, portfolios, short case assessment and long case assessment, log book, trainer's report, audit, simulated patient surgeries, video assessment, simulators, self assessment, peer assessment and standardized patients.

Assessment has a powerful positive steering effect on learning and the curriculum. It conveys what we value as important and acts as the most cogent motivator of student learning. Assessment is purpose driven. In planning and designing assessments, it is essential to recognize the stakes involved in it. The higher the stake, the greater the implications of the outcome of the assessment. The more sophisticated the assessment strategies, the more appropriate they become for feedback and learning.

Measuring progress in acquiring core knowledge and competencies may be a problem if the exams are designed to measure multiple integrated abilities, such as factual knowledge, problem solving, analysis and synthesis of information. Students may advance in one ability and not in another. Therefore, progress tests that are designed to measure growth from the onset of learning until graduation should measure discrete abilities.

Mastery testing (criterion-reflected tests) requires that 100% of the items are measured correctly to determine whether students have attained a mastery level of achievements. In non-mastery testing attainment of 65% of a tested material is considered sufficient.

Global rating scales are measurement tool for quantifying behaviors. Raters use the scale either by directly observing students or by recalling student performance. Raters judge a global domain of ability for example: clinical skills, problem solving, etc

Self assessment (self regulation) is a vital aspect of the lifelong performance of physicians. Self monitoring requires that individuals are able not only to work independently but also to assess their own performance and progress.

Every form of assessment can be used as a self assessment exercise as long as students are provided with 'gold standard' criteria for comparing their own performance against an external reliable measure. Self assessment approaches include: written exams (MCQs, True/False, Essay, MEQs, modified CRQs), performance exams (checklists, global rating, student logbook, portfolio, video, etc).

Oral examination/Viva has poor content validity, higher inter-rater variability and inconsistency in marking. The instrument is prone to biases and is inherently unreliable.

Long Essay Questions can be used for assessment of complex learning situations that can not be assessed by other means (writing skills, ability to present arguments succinctly).

The Short Answer Question (SAQ) is an open ended, semi-structured question format. A structured predetermined marking scheme improves objectivity. The questions can incorporate clinical scenarios. A similar format is also known as Modified Essay Question (MEQ) or Constructed Response Question (CRQ). Equal or higher test reliabilities can be achieved with fewer SEQs as compared to true/false items. If a large amount of knowledge is required to be tested, MCQs should be used. SAQs have a better content coverage as compared to long essay question.

Extended Matching Item is based on a single theme and has a long option list to avoid cueing. It can be used for the assessment of clinical scenarios with less cueing. It is a practical alternative to MCQ while maintaining objectivity and consistency. It can be used in both basic and clinical sciences.

Key Feature Test is a clinical scenario-based paper and pencil test. A description of the problem is followed by a limited number of questions that focus on critical, challenging actions or decisions. It has higher content validity with proper blueprinting.

Long Case involves use of a non-standardised real patient. Long case may provide a unique opportunity to test the physician's tasks and interaction with a real patient. It has poor content validity, is less reliable and lacks consistency. Reproducibility of the score is 0.39; meaning 39% of the variability of the score is due to actual performance of students (signal) and the remaining 61% of the variability is due to errors in measurement (noise) (Norcine,2002). In high stake summative assessment long case should be avoided.

Short Case involves use of three to four non-standardised real patients with one to two examiners. It provides opportunity for assessment with real patients and allows greater sampling than single long case.

Objective Structured Clinical examination (OSCE) consists of multiple stations where each candidate is asked to perform a defined task such as taking a focused history or performing a focused clinical examination of a particular system. A standardized marking scheme specific for each case is used. It is an effective alternative to unstructured short cases.

Mini-Clinical Evaluation Exercise (Mini-CEX) is a rating scale developed by American Board of Internal Medicine to assess six core competencies of residents: medical interviewing skills, physical examination skills, humanistic qualities/professionalism, clinical judgment, counseling skills, organization and efficiency.

Direct Observation of Procedural Skills (DOPS) is a structured rating scale for assessing and providing feedback on practical procedures. The competencies that are commonly assessed include general knowledge about the procedure, informed consent, pre-procedure preparation, analgesia, technical ability, aseptic technique, post-procedure management, and counseling and communication.

Clinical Work Sampling is an in-trainee evaluation method that addresses the issue of system and rater biases by collecting data on observed behaviour at the same time of actual performance and by using multiple observers and occasions.

Checklists are used to capture an observed behaviour or action of a student. Generally rating is by a five to seven point

360-Degree Evaluation/Multisource Assessment consists of measurement tools completed by multiple individuals in a person's sphere of influence. Assessment by peers, other members of the clinical team, and patients can provide insight into trainees' work habits, capacity for team work, and interpersonal sensitivity

In the Logbook students keep a record of the patients seen or procedures performed either in a book or in a computer. It documents the range of patient care and learning experience of students. Logbook is very useful in focusing students on important objectives that must be fulfilled within a specified period of time (Blake, 2001).

Portfolio refers to a collection of one's professional and personal goals, achievements, and methods of achieving these goals. Portfolios demonstrate a trainees' development and technical capacity.

Skill based assessments are designed to measure the knowledge, skills, and judgment required for competency in a given domain.

Test of clinical competence, which allows decisions to be made about medical qualification and fitness to practice, must be designed with respect to key issues including blueprinting, validity, reliability, and standard setting, as well as clarity about their formative or summative function. MCQs, essays, and oral examinations could be used to test factual recall and applied knowledge, but more sophisticated methods are needed to assess clinical performance, including directly observed long and short cases, objective structure clinical examinations, and the use of standardized patients.

The Objective Structure Clinical examination (OSCE) has been widely adopted as a tool to assess students, or doctor's competences in a range of subjects. It measures outcomes and allows very specific feedback.

Other approaches to skill-based assessment include: traditional (Oral exam/viva, long case); alternative formats (tackle the problems associated with traditional orals and long cases by having examiners observe the candidates complete interaction with the patient, training examiners to a structured assessment process, increasing the number of patient problems. Traditional unstructured orals and long cases have largely been discontinued in North America.

While selecting an assessment instrument it is necessary to know precisely what it is that is to be measured. This should reflect course outcomes as different learning outcomes

require the use of different instruments. It is essential to use an instrument that is valid, reliable and feasible (calculating the cost of the assessment, both in terms of resources and time). Full variety of instruments will ensure that the results obtained are a true reflection of the students' performance.

Multiple sampling strategies as the accepted methods used in assessment in clinical competency include OSCE, Short Answer Questions, mini-CEX (Mini Clinical Evaluation Exerciser), Directly Observed Procedural Skills (DOPS), Clinical work sampling (CWS), and 360-degree evaluation.

The assessment is an integral component of overall educational activities. Assessment should be designed prospectively along with learning outcomes. It should be purpose driven. Assessment methods must provide valid and usable data. Methods must yield reliable and generalisable data.

Multiple assessment methods are necessary to capture all or most aspects of clinical competency and any single method is not sufficient to do the job. For knowledge, concepts, application of knowledge ('Knows' and 'Knows How' of Miller's conceptual pyramid for clinical competence) context-based MCQ, extended matching item and short answer questions are appropriate. For 'Shows How' multi-station OSCE is feasible. For performance-based assessment ('does') mini-CEX, DOPS is appropriate. Alternatively clinical work sampling and portfolio or log book may be used.

*Standard setting* involves judgment, reaching consensus, and expressing that consensus as a single score on a test. *Norm Referenced Scores* are suitable for admission exercise that requires selection of a predetermined number of candidates. *Criterion Referenced Standard* (based on predefined test goals and standards in performance during an examination where a certain level of knowledge or skill has been determined as required for passing) is feasible for competency-based examination. Various approaches available include test-centred approach (Agnoff's method and its variations), examinee-centred approach (borderline group method), and several other innovations. *Blueprinting* refers to a process emphasizing that test content should be carefully planned against learning objectives.

The purpose of assessment should direct the choice of instruments. Needs assessment is the starting point of good assessment that identifies the current status of the students before the commencement of the actual educational activities. Needs assessment is used to determine the existing knowledge base, future needs, and priority areas that should be addressed.

Student assessment is a comprehensive decision making process with many important implications beyond the measure of students' success. Student assessment is also related to program evaluation. It provides important data to determine the program effectiveness, improves the teaching program, and helps in developing educational concepts.

Good quality assessment not only satisfies the needs of accreditation but also contributes to student's learning. Assessment methods should match the competencies being learnt and the teaching formats being used.

Competence is a habit of lifelong learning, is contextual (e.g. practice setting, the local prevalence of disease, etc) and developmental (habits of mind and behaviour and practical wisdom are gained through deliberate practice).

### **Further Reading**

- ACGME Outcome Project. Accreditation Council for Graduate Medical Education & American Board of Medical Specialist. Toolbox for assessment methods, version 1.1. [www.acgme.org/outcomes/assess/toolbox.pdf](http://www.acgme.org/outcomes/assess/toolbox.pdf)
- Case S M & Swanson D B. Constructing Written Test for the Basic & Clinical Sciences, 3<sup>rd</sup>. ed. (National Board of Medical Examiners) Philadelphia, PA, USA). 2002. [www.nbme.org/about/itemwriting.asp](http://www.nbme.org/about/itemwriting.asp).
- Day S C, Norcini J J, Diserens D, et al. The validity of the essay test of clinical judgement. *Acad Med.* 1990;65(9):S39-40.
- Epstein RM, and Hundert EM. Defining and assessing clinical competence, *JAMA* 2002;387:226-35.
- Friedman Ben\_David M. Standard setting in student assessment, AMEE education Guide No: 18 (Association for Medical education in Europe, Dundee, UK), 2000.
- Miller GE. The assessment of clinical skills/competencies/performance. *Acad Med.* 1990;65(9):S63-67.
- Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice questions and patient management problem item formats in assessment of clinical competence. *Med Edu.* 1985;19(3):238-47.
- Norman G. Postgraduate assessment – reliability and validity. *Trans J. Coll. Med. S. Afri.* 2003;47:71-75.
- Page G & Bordage G. & Allen T. Developing key feature problem and examination to assess clinical decision making skills. *Acad Med.* 1995;70(3): 194-201.
- Swanson DB. A measurement framework for performance based test. In: Hart IR, Harden RM (eds.), *Further Development in assessing Clinical Competence.* Montreal Can-Heal. 1987.
- Wass, Cees Van der Vleuten, John Shatzer, Roger Jones. Assessment of clinical competence. *The Lancet* 2001;357:945-49.
- Vleuten van der CPM. Validity of final examination in undergraduate medical training. *BMJ* 2000;321:1217-19.
- Falchikov N, Boud D. Student Self-assessment in higher education: a meta-analysis. *Review of Education Research* 1989; S9:345-430.
- Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state-of-the-art teaching and learning in medicine. 1990;22:58-76.

**Syed Amin Tabish**  
FRC, FRCPE, FACP, FAMS, MD  
Editor-in-Chief