

Current Trends of Assessments in Medical Education with Special Reference to Preferred Tools in Anatomy Examinations

Hulail M, Hassan NH, Tharwat M, ElSawy

References

1. Accreditation Council for Graduate Medical Education, American Board of Medical Specialties. Toolbox of assessment methods. <http://www.acgme.org/Outcome/assess/Toolbox.pdf> BMJ 2000;321:1217–19.
2. AERA, APA, NCME. Foundations. standards for educational and psychological testing: Part I. Washington, DC: American Educational Research Association, 2014:11–72
3. Allen D, and Tanner K. “Approaches to cell biology teaching: questions about questions.” Cell Biol. Edu. Vol.1, 2002: 63-67.
4. Ali SK, Bordage G. Validity of key features for a family medicine pilot exam at the College of Physicians and Surgeons Pakistan. J Coll Phys Surg Pakistan. 1995;5(6):256–260.
5. Allen D and Tanner K. “Putting the horse back in front of the cart: using visions and decisions about high-quality learning experiences to drive course design.” CBE Life Sci. Edu. Vol.6, 2007: 85-89.



WJMER

World Journal of Medical Education and Research

An Official Publication of the Education and Research Division of Doctors Academy

Current Trends of Assessments in Medical Education with Special Reference to Preferred Tools in Anatomy Examinations

Knowledge, Attitude and Practices of Diabetic Foot Patients Admitted to the Surgical Wards at Baghdad Teaching Hospital: A Cross-Sectional Study

The Perceived Role of Community-Based Medical Education Among Kenyan-Trained Medical Doctors' Choice of Specialty

Surgical Snapshot: Management of Right Iliac Fossa Pain in Adults and Children

Patient's Autonomy: The Right to Choose Who Patients Consult in a Public Teaching Hospital

Management of Ectopic Pregnancy

The Role of Innate Factors in The Aetiology of Obesity



ISSN 2052-1715



Current Trends of Assessments in Medical Education with Special References to Preferred Tools in Anatomy Examinations

Hulail M, Hassan NH, Tharwat M, ElSawy NA

Institution

44519, Shaibet an
Nakareyah, Zagazig 2, Ash
Sharqia Governorate, Egypt

**WJMER, Vol 23: Issue 1,
2020**

Abstract

The assessment is an essential aspect of medical education. It fosters and motivates student learning and provides them with the skills necessary to develop lifelong learning. It provides the Licensing bodies with the necessary information to certify the future practitioners, and to discriminate among candidates for advanced training. Assessment may be summative to make decisions about the student grade for a course or formative assessment which is a core contributor to learning itself and guides student teaching.

In the first half of the last century, assessment in medical education depended mainly on written and oral examinations. However, traditional assessment tools are poor in measuring skills such as independent learning, communication with patients, working as part of a health team and problem-solving skills. From the fifties of the last century, several new methods of assessment have been developed to measure all aspects of student competence. Multiple-choice questions (MCQs) had been widely used to test knowledge. The objective structured clinical examination (OSCE) was introduced to assess clinical skills. When the objectives of a medical course are changed, the assessment methods should reflect these changes.

All methods of assessment have strengths as well as limitations. The role of an assessment planner is to use different methods to maximize the benefits of such assessment.

The assessment method chosen should align with the nature of the knowledge, skills, or behaviors to be assessed. Medical institutions must strive to produce competent doctors to ensure optimal patient care.

Key Words

Assessments; Education; Anatomy; Examinations.

Corresponding Author:

Mr Mohey Hulail; E-mail: hulailm@yahoo.com

Introduction

Assessment is the measurement of students progress in the educational process. It is an essential aspect of medical education as it fosters and motivates learning of students and helps them to accomplish the skills required for lifelong learning (Snyder, 1971; Bloxham and Boyd, 2007; Epstein, 2007).

If the assessment is conducted properly, it serves multiple purposes (Amin and Khoo, 2004; Newble, 1998). It determines whether the learning objectives are met, it supports the learning and certification processes of students and judges their competency. Moreover, it helps to evaluate the teaching programs to predict future performance. Certifying organizations perform assessments to ensure competence of future practitioners, to select candidates for advanced training, to motivate and direct the learning process and to evaluate the training programs (Epstein and Hundert, 2002).

It outlines the students' experience and behavior more than any other educational tool (Epstein, 2007; Bennett, 2011; O'Neill, 2014). Most students focus on acquiring knowledge and exhibiting skills that they expect to be the assessment requirements (Fransson, 1977).

Vleuten (1996) described that assessment guide the learning process through its frequency, timing, content, design or format regarding what is asked and the knowledge required.

The demands of the society had been changed so medical institutions must teach and assess their students in the most beneficial way to meet these changes (Shaughnessy and Pauline, 2015). It is seen as the single strongest determinant of what students actually learn (as opposed to what they are taught) and is considered to be a uniquely powerful tool for manipulating the whole education process (Stella, 1993).

Harden and Laidlaw (2012) stated that students may neglect bad teaching; however, they cannot behave in the same way with assessments if they seek certifications built on such assessments.

An integrated curriculum that does not incorporate equally integrated assessment strategies is ineffective in achieving the desired educational outcomes (David et al., 2017).

Principles of Assessment

Assessment is necessary of any educational program. The perfectness of an assessment tool is determined by its reliability, validity, fidelity, educational impact, feasibility and acceptability. Careful balanced cooperation between these five principles of assessment is required rather than focus on any one of them. The efficacy of assessment is the product of all five of these characters. For choosing an appropriate assessment tool, this conceptual model had to be considered (Vleuten, 1996; Norcini et al., 2011).

The validity of an assessment tool is the extent to which it measures what it is supposed to measure (Towle, 1991; Goetz et al., 1992; Atkinson et al., 1993; Pressley and McCormick, 1995; Vleuten, 1996; Messick, 1989; Downing, 2003). The assessment must contain a range of components to be assessed as behaviors, attitudes, skills and knowledge. Validity is the interpretation of assessment results rather than the assessment used (Messick, 1989; Downing, 2003).

Reliability is the reproducibility of assessment results over time at various instances. A reliable test should yield reproducible scores or at least a similar ranking of participants if they were retested (Goetz et al., 1992; Atkinson et al., 1993; Pressley and McCormick, 1995). Repetition of the test on the same group of students using the same assessment instrument should give the same results as the earlier scores. This is called the stability of test scores or test-retest reliability. The concept of reliability includes the ability of assessment tools to differentiate between students (ACGME, 2000; Downing 2004; AERA, APA, and NCME 2014).

Feasibility is concerned with the requirements of the assessment itself. Does it require unreasonable resources concerning its time, cost or staff to achieve it? Is the assessment tool is practical to run? Moreover, the cost of an assessment tool is an important criterion in the case of its widespread use in a medical school (Fowell et al., 2000).

Types of Assessments

There are two types of assessments. The first type of assessment is the assessment of learning, also

known as the summative assessment. It is the traditional paradigm in formal education and determines student progression. It provides a summary of a student's learning progress to be used in making decisions about the student's completion of a course or achievement of a grade (Downing and Yudowsky, 2009). It is applied for validation and accreditation of the learner. However, it does not guide future success for both students and educators (Wiliam, 2011).

The second type of assessment is a formative assessment. It is assessment for learning and is a core contributor to learning itself. It guides student teaching (WHO, 2001; Tormey, 2015; Downing and Yudowsky, 2009). It is done in the classroom to evaluate student understanding. Teachers, learners, or their peers can interpret the results of these assessments to decide the next steps in teaching (Broadfoot et al., 2002; Black and Wiliam, 2009). It is performed mainly to enhance student learning by taking feedback on his performance. It is a tool used in medical education to identify the points of strength and weakness in students (Downing and Yudowsky, 2009).

Feedback from formative assessments had a significant effect on summative assessments for both students and teachers. Facing several settings of formative assessment markedly reduced students' fear of summative examinations. It also has an important role in promoting the teaching-learning processes (Begum et al., 2013). Unfortunately, medical educators depend mainly on summative assessment at the end of a course to test acquired knowledge. It becomes too late to correct student deficiencies (Ende, 1983).

Grading System

Different grading systems are used by medical schools. These systems include five-step letter (A, B, C, D, or F) grading based on the following numeric grading system: A Excellent 90 – 100, B Very Good 80 – 89, C Good 70 – 79, D Pass 60- 69, F Fail 59 and below, honors/pass/fail or P/F system (Kim, 2007).

Recently, there has been a tendency to move toward a pass/fail grading system in medical schools in the U.S., particularly in the first two years. This is based on the assumption that it will reduce stress and anxiety among students and improve their psychological status as it diminishes competitiveness and promotes cooperative learning (Robert et al., 2009). In a pass-fail grading system, students will focus on the learning, giving and gaining of feedback to and from their peers on course activities (Nolen, 2011; Kohen, 2011). Rohe et al. (2006) proposed that a pass-fail grading system reflects students'

respect, concern and accountability for their peers and patients. It might have positive consequences in the success of health care reprogramming and reform.

However, other medical schools assumed that a change from a traditional graded system to a pass/fail system may lead to a decline in attendance of scheduled educational activities, academic performance and in United States Medical Licensing Examination (USMLE) Step 1 scores (Robins et al., 1995). There are also reports that the letter grading system is highly reliable and better than the P/F grading system with regard to acquisition of minimal competency requirements in bedside nursing education (Reznick et al., 1989; Andre, 2000). Ravelli and Wolfson (1999) concluded that it is better to categorize grades into four or five marks.

Standard References Setting

External standards are established as a norm and are used to distinguish between students who gain satisfactory knowledge and skills and those who do not have these minimal standards and consequently fail them. However, these standard depend on who settled them and the methods used (Norcini, 2003). However, in norm settled assessments, a predetermined percentage failure is determined. Students' scores in an exam are arranged on a sequential pattern and a pass mark is adjusted to provide the required percentage failure or success rates. Poor students can pass the exam despite not having the desired minimal standards. Conversely, an excellent way of subject teaching may have dramatic effects on the students' knowledge, but not be reflected in any improvement in the previously settled pass rate (Stella, 1993).

Methods of Assessment

Previously, in the first half of the last century, written assessments consisting of essay questions and oral examinations including clinical cases were the main methods of assessment used in medical education (Norcini, 2005).

From the fifties of last century onwards, new assessment tools were introduced to measure students' knowledge, clinical skills and competence. Knowledge was assessed by multiple-choice questions (MCQs). Clinical skills were assessed by structured clinical examination (OSCE). Projects and log books were applied to document practical skills and clinical case examinations (Cushing, 2000).

WHO (2001) advised that traditional methods of assessment are weak in measuring clinical and practical skills, ability of self learning, problem-solving skills, communication with patients or sharing in health team work. Furthermore, methods

of assessment should be obvious and known to students at the beginning of the course and these methods should be changed if the objectives of a medical course are changed to reflect the newly settled changes.

Miller (1990) proposed a four-layered framework termed The Miller Pyramid to assess progression of knowledge and clinical skills. This Pyramid classifies clinical competence into four levels: Knows (at the base of the pyramid); Knows How; Shows How; and Does (at the apex of the pyramid) (Figure 1). During the learning process, the student progresses from 'Knows' at the base of the pyramid to 'Does' at its apex (Downing and Yudowsky, 2009).

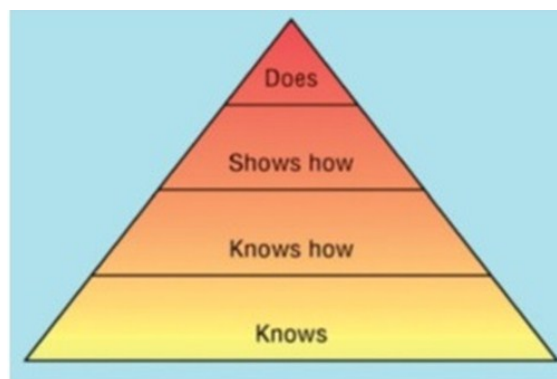


Figure 1: The Miller Pyramid for classification and assessment of clinical competence cited from Downing and Yudowsky (2009)

Assessment of 'Knows' and 'Knows How' of The Miller Pyramid can be done by long and short essay questions, oral examination, multiple-choice questions and extended matching items (EMI). 'Shows How' can be assessed by OSCE, long case and short case. 'Does' at the apex of The Miller Pyramid can be assessed by Mini Clinical Evaluation Exercises (Mini-CEX), Direct Observation of Procedural Skills (DOPS), Checklist, 360-Degree Evaluation, Logbooks and Portfolios (Zubair et al., 2006).

Miller's pyramid simulated a condensed form of Bloom's taxonomy (Bloom, 1984). Its lower two levels correspond to the six categories of the cognitive domain of the taxonomy. These domains in sequence are knowledge, comprehension, application, analysis, synthesis and evaluation/judgement.

Bloom (1956) proposed a theoretical framework; Bloom's Taxonomy, to organize and classify learning objectives and their assessment of higher education. The taxonomy divides these objectives into 3 domains; cognitive (knowledge based), psychomotor (skills) and affective (attitudinal). Bloom described a graded six levels of domains to categorize the

cognitive level. These are remembering, understanding, applying, analyzing, evaluating, and creating (Bloom et al., 1956; Allen and Tanner, 2002, 2007).

These levels resemble a stairway, allowing teachers to promote their students to reach to a higher level of cognition throughout the anatomy course. Moreover, at any advanced level of these domains, a student can also revise the material at a previous one (Ramalingam et al., 2014). Bruce et al. (2015) stated that the level of complexity of assessment can be graded by Bloom's taxonomy (Figure 2).



Figure 2: Bloom's Taxonomy, cited from Bloom et al. (1956)

Oral examination in which one or more examiners ask a student questions in a face-to-face meeting detects the ability of the student to recall and synthesize knowledge (Zubair et al., 2006). It can be used in assessment of critical thinking, application of knowledge and it evaluates self-confidence and assurance (Memon et al., 2010). Its validity, however, depends to a great extent on the examiners (Wenzel and Kirkevang, 2004; Sadaf et al., 2012). Its validity can be increased by selecting the content area with proper design of a marking scheme (Zubair et al., 2006). Lack of standardization of questions leads to high inter-rater variability and marking inconsistency. It is also poor in assessment of higher order knowledge such as problem solving and tends to test factual knowledge (Zubair et al., 2006; Tabish, 2008). Oral assessments or vivas should not be used in high stakes examinations such as judging borderline candidates or in summative assessments owing to its poor reliability (Muzzin and Hart, 1985; Wass et al., 2001).

Multiple-choice questions (MCQs) became widely used in the assessment of both undergraduate and postgraduate medical examination (Shumway and Harden, 2003). They are cost effective, good discriminatory tool and have a high reliability (AIRukban, 2006). They can be used to assess the ability to recall knowledge and clinical skills (Epstein, 2007). They have a large number of included items so can be used to assess a wide range of content.

Consequently, MCQs are suitable for summative examinations so are used in national licensing tests (Epstein, 2007; van der Vleuten, 2000). The multiple choice question is a reproducible tool for assessment of knowledge but it is less valuable in evaluation of clinical skills. However, it is difficult to write perfect MCQs as they require experience and it is time-consuming and costly process (Epstein, 2007; Shumway and Harden, 2003).

Written examination is the most famous assessment tool used in medical education. It has a significant role on how and what students study and learn (Vleuten, 1996). It includes long or short essays questions (Schuwirth and Vleuten, 2004). Long essay questions are used to assess complex learning subjects that are difficult to assess by other tools (Tabish, 2008). They are beneficial in assessing students' ability to summarize and apply known knowledge to new situations (Schuwirth and Vleuten, 2003).

The short essay questions are open ended questions also termed Modified Essay or Constructed Response Questions (CRQ). They have a better coverage area compared to long essay questions and can be used in the assessment of clinical skills (Tabish, 2008). They facilitate assessment of student creativity and critical thinking. The student has to write the required answers rather than to select it as in MCQs. However, they are more time-consuming than MCQs and have a lower reliability, making them unsuitable for broad sampling (Schuwirth and Vleuten, 2003).

The Modified Essay Question (MEQ) is suitable for the assessment of one's problem-solving ability as the latter is one of the most difficult skills to be properly evaluated (Marshall, 1977; Rabinowitz, 1987). An essay carries the risk of considerable variation in marking and is time-consuming to mark (Wood, 2003). Multiple markers can be used to mark essay questions either short or long to gain good accuracy (Downing, 1992). A detailed answer key for the person marking the paper should be included in essay questions (Gronlund, 2006).

Two methods of scoring are used when marking essay questions. These are the analytic (point-scoring) or global scoring methods. The analytic method uses a model answer that is formerly prepared and the answers required are divided into several parts with marks assigned to each part. The use of multiple markers improves the reliability. Unlike the analytic method, global scoring requires the examiner to read the whole essay and make a judgment about its quality in the form of a letter (e.g. A to E) or Likert scale-type (e.g. fail, borderline fail, fair, good, excellent). The analytic method is

more useful in focused essay questions (Rita and Singh, 2012).

Key feature questions were firstly suggested by Bordage and Page (1987) to assess practical skills and patient solving problems in clinical exams instead of written assessments (Bordage and Page, 1987). A clinical problem is presented followed by several questions that require a written answer or by selecting the best of several possible choices.

It resembles real patient problem and facilitates learning of clinical skills and their related clinical case-based solving. It is used to assess both undergraduate and post-graduate students particularly in high stakes examinations as it helps stakeholders to evaluate problem solving ability, skills of decision-making as well as practical procedural knowledge. If properly constructed it is feasible, highly reliable and has a broad content validity (Page et al., 1995; Ali and Bordage, 1995; Trudel, 2008). Health educators are interested by key features questions over the last decade (Farmer and Page, 2005). It is widely used internationally in assessment medical education (Bordage and Page, 1987; Khan and Aljarallah, 2011; Wallerstedt et al., 2012).

Anatomy assessment despite integrated, could accurately evaluate the students' anatomy knowledge if the assessment tasks are planned for a deep learning approach (Logan and Marskak, 2011). Therefore, direct questions that aim at knowledge regurgitation, students are more associated with surface approach of learning. However, questions designed to conduct and organise knowledge (relational thinking), before applying and synthesising that knowledge into context (elaborated thinking), students in these cases are more likely to adopt deep approaches to learning (Smith et al., 2014; Vorstenbosch et al., 2014)

Extended Matching questions can be used to evaluate clinical scenarios provided that long option list are constructed to avoid cueing. It is a suitable assessment tool that can replace MCQ in basic and clinical sciences as it has less cueing (Tabish, 2008).

True or false questions are easy to write and can assess a broad areas of knowledge however it is difficult to construct properly. Also, they are poor as a discriminatory tool between different students. Their use in medical education becomes questioned (Chandratilake et al., 2011).

The Objective Structured Clinical Examination (OSCE) is an assessment tool firstly described by Harden et al. (1975). It is composed of several stations to assess student performance, clinical and

practical skills. The stations may include real or standardized patients, simulators of procedures, answering clinical notes or interpreting laboratory data in writing format (Harden and Gleeson, 1979; Collins and Harden, 1998; Downing, 2004; Downing and Yudkowsky, 2009). It could be used to assess clinical skills as patient history-taking, knowledge depth, ability to search for a diagnosis and designing management plan. Communication with the patients and their relatives also could be evaluated (Regehr et al., 1998; Swartz et al., 1999).

Students pass through several stations changing from station to another in sequence (Harden and Gleeson, 1979). The duration of each station ranges from five to thirty minutes according to the type of the exam. Assessment of wide range of skills requires short-timed station. While examining clinical skills or patient-case solving requires relatively long-timed stations (Epstein, 2007; Downing and Yudkowsky, 2009).

The reliability of OSCE depends on its wide sampling (Reznick et al., 1993; Norman, 2002). The greater the number of stations in the OSCE, the greater its reliability and content validity are (Norman, 2002). A greater number of stations permit a wide sampling of the clinical and patient care skills to be assessed. The student has to pass through a minimum of 10 stations over the course of 3–4 hours to achieve minimal reliability and generalizability in OSCE stations (Downing, 2004; Downing and Yudkowsky, 2009).

It is accepted that an OSCE with about 20 stations allows the minimal required reliability (ACGME, 2000; Van der Vleuten and Swanson, 1990). If the time for its station is too short its validity might be lost (Wass et al., 2001).

Participation of students in an OSCE examination promotes their learning ability and clinical skills acquisition helping them at the subsequent exams (Carraccio and Englander, 2000).

The disadvantages of OSCE format include that its preparing is expensive and costly (Cusimano et al., 1994). Using real patients is difficult and disturbing as history taking is repeatedly done by the same asked questions (Yudkowsky, 2002). It has no benefit in the assessment of practical skills, management and follow-up of patient care or dealing with life-threatening conditions (Shannon and Norman, 1995).

Long case assessment is composed of examination of a real patient by the student at about 30–45 minutes uninterrupted and unobserved. The student presents his/her observed findings to the examiners

who examine the student orally about the patient case and the management plane. Traditionally students are examined in a single case by one or two examiners. It should not be used in summative assessment especially in high stake situations as its generalizability is limited and the reproducibility of its score is 0.39 meaning that actual student performance is responsible for 39% of the variability, however measurement error accounts for 61% of the score variability (Norcini, 2002). However variations among long clinical cases are reflected in students' scores more than variations between the examiners themselves (Wilkinson, 2008). For these reasons assessment in long cases become largely prohibited in western countries (Norcini, 2002; Tabish, 2014).

Short Case includes assessment in 3 or 4 real clinical cases by one to two examiners (Norcini, 2002). They promote discrimination between students' performance either good or poor than in long cases (Hijazi et al., 2002).

It gives the opportunity of assessment with real patients and has a wider sampling range than the assessment in single long case. It enhances assessment of clinical skills. Its validity is good. However, its reliability is poor and associated with inter-rater variability (Norcini, 2002; Zubair et al., 2006).

Mini Clinical Examination (Mini-CEX) is an assessment tool can be carried out in any healthcare center. An examiner evaluates a student-patient meeting. This meeting is about 15 minutes during which the student is expected to perform history-taking and clinical examination and then presents her/his diagnosis and treatment plan to the examiner (Norcini, 2003). The student performance is marked by a structured scheme and educational feedback is provided. Mini-CEX is a useful tool for assessment of residents in workplace. Six patient-meetings assessed by different examiner for each meeting during the year representing from different clinical problem well chosen from a list of clinical problems (Norcini, 2003). However, it has the same reliability as structured examinations using standardized patients (Vleuten et al., 2010). Four Mini-CEX assessments in the same context are adequate to achieve sufficient reliability. However, it may be difficult to get faculty members to accomplish that in one year. Another disadvantage of the Mini-CEX is that the observations are task- and content-specific (Norcini et al., 2003).

DOPS is a variety of the mini-CEX in which an examiner demonstrates the student performance while doing a practical procedure as venipuncture,

giving an intravenous or intramuscular injection or measuring the blood pressure. The examiner then evaluates the student performance and provides feedback. Six observed encounters as in mini-CEX should be done by the student at the year. Each of these encounters is selected from a previously approved list and done with a different assessor. Generally, four to eight encounters give a satisfactory evidence regarding student performance (Norcini and Mckinley, 2007).

The portfolio is a framework that contains an evidence of accomplished learning outcomes over time. It used by educational organizations to document student learning progress for both formative and summative assessment (Davis et al., 2001). Royal College of General Practitioners, it introduced this tool of assessment in 1993 (Snadden, 1999; Buckley et al., 2009). It reflects what has been learned (Rees, 2005). It typically contains written documents, video or audio recordings, photographs and multimedia, and can be maintained in an electronic format. Its contents included personal experiences, learning process, documentation, certificates, and future goals regarding learning and its environments. Portfolio is closely related to self-directed learning and is most beneficial in evaluating gained competences such as improvement of practical skills and scientific base and profession in patient care (Snadden, 1999; Buckley et al., 2009).

Reliability of portfolios depends on well settled standards and criteria for its contents. Also, it is enhanced by incorporating evidence from a number of sources (Friedman et al., 2001).

Logbooks, like portfolios, document the student's experiences. However, they are usually more limited in scope than portfolios and are focused on data collected in a specific area or activity. At least three kinds of logs have been documented; procedural, operative and case logs (ACGME, 2000). Procedural logs usually document how many and when procedures were performed by the learner. Operative logs are similar but document what was done and when. Logbooks are limited in their assessment powers (ACGME, 2000).

Checklist Evaluation

Checklists include the desired behavior, practical activities or clinical skills that are required to be main characters of learner attitude. Typical criteria include completeness of the work and proper follow-up for correction of the mistakes occurred while performing it. Standards of performance are required to be settled as a guide for evaluation of optimal performance (ACGME, 2000).

Multisource or 360-Degree Assessments

These are measurement tools performed by several people in a person's work habits, capacity for teamwork, interpersonal behaviors (Ramsey et al., 1993; Violato et al., 1997). Evaluators include personal peers, clinical team members or patients. To be effective multisource feedback must give feedback including descriptive comments, statistical data, good monitoring and follow-up. This tool of assessments has a good impact on future academic and clinical performance. However, it is challenging to conduct it with large numbers of evaluators to gather reports regarding students/resident's performance though it may be feasible using an electronic system (Norcini, 2003).

Choosing Assessment Method

Traditional medical education focused on acquiring essential knowledge on each topic. Recently more attention is given to the practical and clinical skills and attitudes in a future good doctor (GMCC, 1980; Towle, 1991). The traditional summative assessment has a great benefit in the early years of medical school as basic knowledge is the core of the curriculum, there are different methods of assessment in the anatomy curriculum. They range from written examinations to practical evaluations such as spotter tests and oral-based assessments such as viva voce (Vorstenbosch et al., 2014). The written component of an anatomy examination usually involves MCQs (consisting of one stem and up to five distractors with one best answer), EMQs (consisting of short cases called vignettes (Wood, 2003), a number of distractors and one best answer for each stem) or short-answer questions (SAQs). However, in the later years medical educators try to promote critical thinking and clinical skills required for future physician the traditional summative assessment is lacking. Medical schools had incorporated more formative forms of assessment to shift from wrote learning to high quality more engaging integrated learning (Ferris and Flynn, 2015).

There is no single assessment tool can assess the wide range of knowledge and skills required for the future physician. Medical examiners should identify the aspects that they wish to test and then provide a range of appropriate tools (Roberts and Norman, 1990). All assessment tools have strengths in some aspects and limitations in others. The assessment manager has to use different tools to gain a maximal benefit from such an assessment (Epstein, 2007). The assessment planner should address six key questions; why is the assessment is to be done, what subjects are to be assessed, how and by any tool, when, where and who (Harden and Laidlaw, 2012). Three major characters of assessment must

be taken in mind during choosing an assessment tool. Firstly, the content of the assessment should be proportionally represented and organized a process known as blueprinting (Raymond and Neustel, 2006). Secondly the validity, reliability of the assessment method to be used. Thirdly the circumstances at which the assessment will be carried out (Kern et al., 1998; Norman, 2006).

The chosen assessment tool must coincide perfectly with the type of knowledge, nature of skills, or behaviors to be assessed (Norman, 2006).

MCQs, EMQs and essay questions can assess and are beneficial in assessment of knowledge. Traditional methods including assessment in long and short cases can be used in assessment of practical and clinical skill. Moreover, newer methods as OSCE (objective-structured clinical examinations) mini-CEX and portfolios can be used in assessment and evaluation of practical and clinical skills (Sood and Singh, 2012). Competency as professionalism, communication skills, attitude, teamwork ability are difficult to evaluate (Epstein, 2007; Roberts and Norman, 1990).

The test content is organized by the blueprinting process through which the test questions have to proportionally represent the content areas of knowledge or behaviors to be tested (Linn, 2006). Though it is difficult to assess everything using a carefully constructed blueprint ensures sufficient and appropriate sampling of all knowledge and skills students are expected to know (Hamdy, 2006).

In the traditional curriculum, anatomy had its own place in assessment, and students had to sit through all three forms of assessment (written, practical and oral). In fact, prior to 2005, medical students were assessed more often by practical (94.2% v. 33.3%) and oral (84.5% v. 13.1%) methods and less often by written methods (25.2% v. 68.7%) when compared with graduates' post 2005 (Rowland et al., 2011). However, despite being used for a considerable amount of time in medical education, the practical spotter tests have been criticized for testing low levels of knowledge in students (Yaqinuddin et al., 2013). Thus, in the modern curriculum, spotter examinations have been eliminated in most institutions (Smith and McManus, 2015).

Although anatomy assessment in the modern curriculum is integrated, it can still function as a reliable tool for demonstrating students' anatomy knowledge if the assessment tasks are designed to foster a deep approach to learning (Logan and Marskak, 2011). Hence, if questions are primarily aimed at knowledge regurgitation, students are

more likely to employ surface approaches to learning.

Defining the purpose of the assessment either high or low stakes help to determine the number of the samples needed. The higher the stakes, the more samples are required in the assessment (Downing, 2003).

Challenges of Assessments

MCQs are difficult to write and can result in cueing, essay questions can have high interrater reliability, oral examinations are subjective and may have sex and race unfairness and simulators are expensive and present artificial settings. To overcome these challenges, multimodal assessment can be used to provide broader insights into trainee competence and allows input from a number of assessors of the student's performance (Holmboe et al., 2010).

Educators tend to assess what is easy, rather than assessing what is important. On the other hand, excessive variability in ratings by different raters has raised concerns about reliability and validity (Gingerich et al., 2011). All measurements have error (Downing, 2003).

The quality of patient care is directly linked with a clinician's attitude (CQHCA, 2001). Poor academic quality often results in deficient clinical care despite the underlying individual's level of experience (Hamilton and Pinnegar, 2000). Medical institutions must strive to produce competent doctors to ensure optimal patient care (Shaughnessy and Pauline, 2015). However, the assessment in medicine is greatly challenging due to combination of increasing cohort sizes and shrinking budgets (Fottrell, 2006; Gibbs, 2006).

Recently medical migration occurred worldwide with greater tendency towards more rich and comfortable countries. Comparison of students and clinicians from different medical educational backgrounds to select for training posts or employee became more difficult (Bidwell et al., 2013; Chen and Boufford, 2005).

Conclusion

Assessment is of two types. Summative assessment determines student progression and to provide a summary of a student's performance to be used to make decisions about the student grade and validation and accreditation purposes. Formative assessment is a core contributor to learning itself. It guides student learning and gives an evidence of where the student is at the course and using this evidence can be used by teacher to make decisions about the next steps in instruction and had a significant effect on summative assessment in

various aspects. No single examination tool can be used to assess the wide range of knowledge and skills required for the future physician. Medical examiners should identify the aspects that they wish to test and then provide a range of appropriate tools. The content of the assessment should be proportionally represented and organized a process known as blueprinting. The chosen assessment tool must coincide perfectly with the type of knowledge, nature of skills, or behaviors to be assessed. MCQs, EMQs and essay questions can assess and are beneficial in assessment of knowledge. Traditional methods including assessment in long and short cases can be used in assessment of practical and clinical skill. Moreover, newer methods as OSCE (objective-structured clinical examinations) mini-CEX and portfolios can be used in assessment and evaluation of practical and clinical skills. Competency as professionalism, communication skills, attitude, teamwork ability are difficult to evaluate. The test content is organized by the blueprinting process through which the test questions have to proportionally represent the content areas of knowledge or behaviors to be tested.

Acknowledgements

None.

Conflicts of Interest

The authors declare there is no conflicts of interest.

References

1. Accreditation Council for Graduate Medical Education, American Board of Medical Specialties. Toolbox of assessment methods. <http://www.acgme.org/Outcome/assessment/Toolbox.pdf> BMJ 2000;321:1217-19.
2. AERA, APA, NCME. Foundations. standards for educational and psychological testing: Part I. Washington, DC: American Educational Research Association, 2014:11-72
3. Allen D, and Tanner K. "Approaches to cell biology teaching: questions about questions." Cell Biol. Edu. Vol.1, 2002: 63-67.
4. Ali SK, Bordage G. Validity of key features for a family medicine pilot exam at the College of Physicians and Surgeons Pakistan. J Coll Phys Surg Pakistan. 1995;5(6):256-260.
5. Allen D and Tanner K. "Putting the horse back in front of the cart: using visions and decisions about high-quality learning experiences to drive course design." CBE Life Sci. Edu. Vol.6, 2007: 85-89.
6. Amin Z, Khoo HE, Tan CH. A novel approach to faculty development programme evaluation. Med Educ. 2004. 38(11):1187-8. doi:10.1111/j.1365-2929.2004.01991.x PMID:15507020.

7. Andre K. Grading student clinical practice performance: the Australian perspective. *Nurse Educ Today*. 2000;20:672-9.
8. Atkinson RL, Atkinson RC, Smith EE and Bem DJ. *Introduction to Psychology*, 1993, 11th edn (New York, Harcourt Brace Jovanovich College Publishers).
9. Begum N, Hossain S and Talukder HK. Influence of formative assessment on summative assessment in undergraduate medical students Bangladesh Journal of Medical Education Vol.- 04, Issue- 01, 2013
10. Bennett RE. Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 2011: 5-25. <http://dx.doi.org/10.1080/0969594X.2010.513678>
11. Bidwell P, Humphries N, Dicker P, Thomas S, Normand C, and Brugha R. The national and international implications of a decade of doctor migration in the Irish context. *Health policy*, 110(1), 2013: 29-38. <http://dx.doi.org/10.1016/j.healthpol.2012.10.002>
12. Black P, Wiliam D. Developing a theory of formative assessment. *Educ Assess Eval Account* 2009, 21:5–31
13. Bloom BS. Taxonomy of educational objectives. Handbook I: Cognitive domain. New York, 1956
14. David McKay, Bloom B. Taxonomy of Educational Objectives. In: D. McKay, ed., *The Cognitive Domain*, 1984: 1st ed. New York: Company Inc.
15. Bloxham S, and Boyd P. *Developing Effective Assessment In Higher Education: A Practical Guide: A Practical Guide*. 2007, McGraw-Hill International.
16. Bordage G, Page G. An alternative approach to PMPs: The "key features" concept. In: Hart IR, Harden RM (Hrsg). *Further developments in assessing clinical competence*. Montreal: Can-Heal; 1987. S.59–75.
17. Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. *Assessment for learning: 10 principles*. University of Cambridge School of Education. <http://arrrts.gtcni.org.uk/gtcni/bitstream/2428/4624/1/Testing,%20Motivation%20and%20Learning.pdf>. Published 2002.
18. Buckley S, Coleman J, Davison I, Khan KS, Zamora J, Malick S, Moreley D, Pollard D, Ashcroft T, Popovic C, Sayers J. The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11. *Med Teach*. 2009;31(4):282–298. DOI: 10.1080/01421590902889897
19. CARRACCIO C. and ENGLANDER R. The objective structured clinical examination: a step in the direction of competency-based evaluation, *Archives of Pediatric Adolescent Medicine*, 154, pp. 2000: 736–741.
20. Chandratilake M, Davis M, Ponnampuruma G. Assessment of medical knowledge: The pros and cons of using true/false multiple-choice questions. *Natl Med J India* 2011; 24:225–8.
21. Chen LC and Boufford JL. Fatal flows—doctors on the move. *New England Journal of Medicine*, 353(17), 2005: 1850–1852. <http://dx.doi.org/10.1056/NEJMe058188>
22. Cohen L, Manion L and Morrison K. *Research Methods in Education* (7th ed.) 2011, London & New York: Routledge.
23. Collins J, Harden RM. AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. *Med Teach*. 1998;20(6):508–521. DOI: 10.1080/01421599880210
24. Committee on Quality of Health Care in America-CQHCA. *Crossing the quality chasm: a new health system for the 21st century*. 2001, Washington, DC: National Academy Press.
25. CUSHING A. Developments in attitude and professional behavior assessment, oral presentation given at the 9th International Ottawa Conference on Medical Education, 2000, Cape Town, South Africa, 28 February–3 March [<http://www.educ.unimaas.nl/ottawa/>].
26. Cusimano MD, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R. A comparative analysis of the costs of administration of an OSCE (objective structured clinical examination). *Acad Med* 1994; 69:571–6.
27. David W. Fielding, EdD, a Glenn Regehr, PhD A Call for an Integrated Program of Assessment. *American Journal of Pharmaceutical Education* 2017; 81 (4).
28. Davis MH, Friedman Ben-David M, Harden RM, Howie P, Ker J, McGhee C, et al. 364 THE NATIONAL MEDICAL JOURNAL OF INDIA VOL. 25, NO. 6, 2012 Portfolio assessment in medical students' final examinations. *Med Teach* 2001;23: 357–66.
29. Downing S, Yudkowsky R, editors. *Assessment in Health Professions Education*. New York, NY: Routledge; 2009.
30. Downing SM. Reliability: On the reproducibility of assessment data. *Medical Education* 2004; 38:1006-1012.
31. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830–837.
32. Downing SM: True-false, alternate-choice, and multiple choice items. *Educ meas, issues pract* 1992, 11:27-30.
33. Ende J. Feedback in clinical medical education. *JAMA* 1983;250:777-81.

34. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002 Jan 9;287 (2):226-35.
35. Epstein RM. Assessment in medical education. *New England Journal of Medicine*, 356(4), 2007: 387-396. <http://dx.doi.org/10.1056/NEJMr054784>
36. Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 2005;39:1188-94.
37. Ferris, H & Flynn, D.O Assessment in Medical Education; What Are We Trying to Achieve? *International Journal of Higher Education* Vol. 4, No. 2; 2015
38. Fottrell. (2006). Report of the Working Group on Undergraduate Medical Education and Training. <http://health.gov.ie/wp-content/uploads/2014/05/fottrell.pdf>.
39. Fowell S, Maudsley G, Maguire P, Leinster S and Bligh J. Student assessment in undergraduate medical education in the United Kingdom 1998. *Medical Education*, 2000, 34: Suppl. 1,
40. Fransson A. 'On qualitative differences in learning. IV – effects of intrinsic motivation and extrinsic test anxiety on process and outcome', *British Journal of Educational Psychology* 47, 1977: 244–257.
41. Friedman Ben David M, Davis M, Harden R, Howie P, Ker J, Pippard M. AMEE guide No. 24. Portfolios as a method of student assessment. *Medical Teacher* 2001; 23:535-551.
42. General Medical Council Education Committee. Recommendations on basic medical education. London: (GMCC 1980, Towle 1991).
43. Gibbs G. Why assessment is changing. *Innovative assessment in Higher Education*, 11, 2006.
44. Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: rethinking the etiology of rater errors. *Acad Med*. 2011; 86 (suppl 10):1–7.
45. GOETZ ET, ALEXANDER PA and ASH MJ. *Educational Psychology A Classroom Perspective* (New York, Maxwell Macmillan International), 1992.
46. Gronlund NE (ed). *Assessment of student achievement*. 8th ed. Boston, MA:Pearson Education; 2006:3–13.
47. Hamdy H. Blueprinting for the assessment of healthcare professionals. *Clin Teach*. 2006;3 (3):175–179.
48. Hamilton ML and Pinnegar S. On the threshold of a new century trustworthiness, integrity, and self-study in teacher education. *Journal of Teacher Education*, 51(3), 2000: 234-240. <http://dx.doi.org/10.1177/0022487100051003012>
49. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *J Med Educ* 1979;13:41–54.
50. Harden RM and Laidlaw JM. *Essential skills for a medical teacher: an introduction to teaching and learning in medicine*. Elsevier Health Sciences, 2012.
51. Hijazi Z, Premadasa IG and Moussa MA. Performances of students in the final examination in paediatrics: importance of short cases, *Arch. Dis. Childhood* 2002, 86: 57-58. 20.
52. Holmboe ES, Sherbino J, Long DM, Swing SR and Frank JR. The role of assessment in competency-based medical education. *Medical Teacher*, 2010, 32(8), 676-682. <http://dx.doi.org/10.3109/0142159X.2010.500704>
53. Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL, Scalese RJ. Features and uses of high fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher* 2005; 27:10-28.
54. Kern DE, Thomas PA, Howard DM, Bass EB. *Curriculum Development for Medical Education: A Six Step Approach*. Baltimore and London: Johns Hopkins University Press; 1998.
55. Khan MU, Aljarallah BM. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. *Int J Health Sci*. 2011;5(1):39–43.
56. Kim M Y Is the Pass/Fail System Applicable to a Medical School in Korea? *J Educ Eval Health Prof* 2007, 4: 3 DOI 10.3352/jeehp. 2007. 4. 3
57. Linn RL. The standards for educational and psychological testing: Guidance in test development. In: Downing SM, Haladyna TM, ed. *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates; 2006.
58. Logan J. & Marskak D. Testing to enhance retention in human anatomy. *Anatomical Sciences Education*, 2011; 4, 243–248.
59. Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Adv Health Sci Educ*. 2010;15(2):277–289. DOI: 10.1007/s10459-008-9111-9
60. Messick S. (1989): Validity. In: *Educational Measurement*, R, L. Linn (ed) 3 rd ed. McMillan New York. 1989. pp. 13-103.
61. Miller GE. The assessment of clinical skills/ competence/ performance. *Acad Med* 1990, 65:563–567
62. Muzzin LR and Hart L. Oral Examinations. In: Neufeld, V.R. and Norman, G.R., eds. 1985 *Assessing Clinical Competence*. New York, Springer Publishing Company, 71-93. As reviewed in Fowell, S., Maudsley, G., Maguire, P., Leinster, S. and Bligh, J. 2000. Student

- assessment in undergraduate medical education in the United Kingdom 1998. Medical Education, 34: Suppl. 1, 1–49. doi: 10.1046/j.1365-2923.2000.0340s1001.x
63. Newble DI. Assessing clinical competence at the undergraduate level. Med Educ 1992;26:504–11.
64. Nolen SB. The role of educational systems in the link between formative assessment and motivation. Theory Into Practice. 2011;50 (4) : 3 1 9 - 3 2 6 . d o i : 0.1080/00405841.2011.607399.
65. Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206 2 1 4 . h t t p s : / / doi.org/10.3109/0142159X.2011.551559
66. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. Ann Intern Med. 2003;138(6):476–481. DOI: 10.7326/0003-4819-138-6-200303180-00012
67. Norcini JJ, Blank LL, Duffy FD and Fortna GS. The mini-CEX: a method for assessing clinical skills. Ann Intern Med. 2003, 138(6):476–481. DOI: 10.7326/0003-4819-138-6- 200303180-00012
68. Norcini JJ, Mckinley DW. Assessment methods in medical education. Teach Teacher Education, 2007: 23, 239-250
69. Norcini JJ. The death of the long case? BMJ. 2002; 324(7334):408–409. DOI: 10.1136/bmj.324.7334.408
70. Norcini JJ. The mini clinical evaluation exercise (mini-CEX). ClinTeach. 2005;2(1):25-30. DOI: 10.1111/j.1743- 498X.2005.00060.x
71. Norman G. How specific is case specificity? Medical Education. 2006; 40:618-623.
72. Norman G. The long case versus objective structured clinical examinations. BMJ 2002; 324:748-9.
73. O'Neill G. Purposes & Principles of Assessment in Higher Education. University College Dublin. 2014, <http://www.ucd.ie/teaching/resources/assessment/presentation/presentation.html>.
74. Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. Acad Med. 1995;70(3):194–201. DOI: 10.1097/00001888-199503000-00009
75. Pressley M and McCormick CB. Advanced Educational Psychology for Educators, Researchers, and Policymakers (New York, HarperCollins College Publishers), 1995.
76. Rabinowitz HK. The modified essay question: an evaluation of its use in a family medicine clerkship. Medical Education 1987, 21:114-18.
77. Ramalingam, Kasilingam, Mritha ramalingam, Elanchezian Chinnavan. Assessment of learning domains to improve student's learning in higher education Journal of Young Pharmacists Vol 6, Issue 4, Jan-Mar, 2014
78. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. JAMA 1993;269:1655–60.
79. Ravelli C, Wolfson P. What is the “ideal” grading system for the junior surgery clerkship? Am J Surg. 1999;177:140-4.
80. Raymnod M, Neustel S. Determining the content of credentialing examinations. In: Downing SM, Halyadyna TM, ed. Handbook of Test Development. Mahwah, NJ: Lawrence Erlbaum Associates; 2006.
81. Rees C. The use (and abuse) of the term ‘portfolio’ Med Edu 2005;39:436–7.
82. Regehr G, MacRae H, Reznick R, 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.
83. Reznick RK, Blackmore D, Cohen R, Baumber J, Rothman A, Smee S, et al. An objective structured clinical examination for the licentiate of the Medical Council of Canada: From research to reality. Acad Med 1993;68 (Suppl):S4–S6.
84. Rita S and Singh T. Assessment in medical education: Evolving perspectives and contemporary trends. National Medical Journal Of India VOL. 25, NO. 6, 2012.
85. Robert A. Bloodgood, PhD, Jerry G. Short, PhD, John M. Jackson, MS, and James R. Martindale, PhD A Change to Pass/Fail Grading in the First Two Years at One Medical School Results in Improved Psychological Well-Being Academic Medicine, Vol. 84, No. 5 / May 2009
86. Roberts J and Norman G. Reliability and learning from the objective structured clinical examination. Med Educ 1990;24:219-23.
87. Robins LS, Fantone JC, Oh MS, Alexander GL, Shlafer M, Davis WK. The effect of pass/ fail grading and weekly quizzes on first-year students' performances and satisfaction. Acad Med. 1995;70:327–329.
88. Rohe DE, Barrier PA, Clark MM, Cook DA, Vickers KS, Decker PA. The benefits of pass-fail grading on stress, mood, and group cohesion in medical students. Mayo Clin Proc. 2006;81:1443 –8.
89. Rowland S, Ahmed K, Davies DC, Ashrafian H, Patel V, Darzi A, Paraskeva PA, Athanasiou T. Assessment of anatomical knowledge for clinical practice: Perceptions of clinicians and students. Surg Radiol Anat, 2011; 33:263–269.
90. Sadaf S, Khan S, Ali SK. Tips for developing a valid and reliable bank of multiple choice

- questions (MCQs). *Educ Health*. 2012;25(3):195–197. DOI: 10.4103/1357-6283.109786
91. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* 2004;38:974-9.
 92. Schuwirth LWT and Vleuten CPM. Written assessment. In: Cantillon P, Hutchison L, and Wood D (eds.) *ABC of Learning and Teaching in Medicine*, 2003, 29-31 (BMJ Publishing Group, UK
 93. Shannon S, Norman G (eds). *Evaluation methods: A resource handbook*. Hamilton, Ont: McMaster University Program for educational development; 1995:71–7.
 94. Shaughnessy SMO and Pauline J. Summative and Formative Assessment in Medicine: The Experience of an Anaesthesia Trainee. *International Journal of Higher Education* Vol. 4, No. 2; 2015
 95. Smith, C. F., & McManus, B. The integrated anatomy practical paper: A robust assessment method for anatomy education today. *Anatomical Sciences Education*, 2015; 8, 63–73.
 96. Smith, C. F., Martinez-Alvarez, C., & McHanwell, S. (2014). The context of learning
 97. anatomy: Does it make a difference? *Journal of Anatomy*, 224(3), 270–278.
 98. Snadden D. Portfolios: attempting to measure the unmeasurable? *Medical Education* 1999; 33:478-479.
 99. Snyder BR. *The hidden curriculum*. New York: Knopf, 1971.
 100. SOOD R and SINGH T. Assessment in medical education: Evolving perspectives and contemporary trends *THE NATIONAL MEDICAL JOURNAL OF INDIA* VOL. 25, NO. 6, 2012
 101. Stella Lowry: *Medical education. Assessment of students* BMJ volume 306 1993
 102. Swartz MH, Colliver JA, Bardes CL, Charon R, Fried ED, Moroff S. Global rating of videotaped performance versus global ratings of actions recorded on checklists: A criterion for performance assessment with standardized patients. *Acad Med* 1999;74:1028–3256
 103. Tabish SA. Assessment methods in Medical Education. *Int J Health Sci (Qassim)*; 2008; 2 (2):3-7
 104. Tabish SA. *Assessment Methods in Medical Education*, 2014. [https:// www.researchgate.net/ publication/240469755](https://www.researchgate.net/publication/240469755)
 105. Tormey W. Education, learning and assessment: current trends and best practice for medical educators *Ir J Med Sci* (2015) 184:1–12
 106. Towle A. *Critical thinking: the future of undergraduate medical education*. London: King's Fund Centre, 1991.
 107. Trudel JL, Bordage G, Downing SM. Reliability and validity of key feature cases for the self-assessment of colon and rectal surgeons. *Ann Surg*. 2008;248(2):252–258. DOI:10.1097/SLA.0b013e31818233d3
 108. van der Vleuten CM, Swanson DB. Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine* 1990; 2:58-76.
 109. van der Vleuten CP, Verwijnen GM, Wijnen W. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach*. 1996;18(2):103–109. DOI: 10.3109/01421599609034142
 110. van der Vleuten et al. *Best Practices Res Clin OBS/GYN* 2010, p. 712
 111. Violato C, Marini A, Toews J, Lockyer J, Fidler H. Feasibility and psychometric properties of using peers, consulting physicians, co-workers, and patients to assess physicians. *Acad Med* 1997;72 (10 Suppl 1):S82–S84.
 112. Vorstenbosch MA, Bouter ST, Van der Hurk MM, Kooloos JG, Bolhuis M & Laan RF. Exploring the validity of assessment in anatomy: Do images influence cognitive processes used in answering extended matching questions? *Anatomical Science Education*, 2014, 7(2), 107-116.
 113. Wallerstedt S, Erickson G, Wallerstedt SM. Short Answer Questions or Modified Essay questions—More Than a Technical Issue. *Int J Clin Med*. 2012;3:28. DOI: 10.4236/ijcm.2012.31005
 114. Wass C, van der Vleuten C. The long case. *Med Educ* 2004;38:1176–80.
 115. Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945–949. DOI: 10.1016/S0140-6736(00)04221-5
 116. Wenzel A, Kirkevang L. Students' attitudes to digital radiography and measurement accuracy of two digital systems in connection with root canal treatment. *Eur J Dent Educ*. 2004;8(4):167–171. DOI: 10.1111/j.1600-0579.2004.00347.x
 117. WHO Guidelines for Quality Assurance of Basic Medical Education in the Western Pacific Region *WORLD HEALTH ORGANIZATION REGIONAL OFFICE FOR THE WESTERN PACIFIC* Manila, Philippines July 2001
 118. Wiliam, D. What is assessment for learning?. *Studies in Educational Evaluation*. 2011; 37(1): 3–14. doi:10.1016/j.stueduc.2011.03.001.
 119. Wilkinson TJ, Campbell PJ, Judd SJ. Reliability of the long case. *Med Educ* 2008;42:887–93.

120. Wood EJ. What are extended Matching Sets Questions? Bioscience Education eJournal 2003, 1: [http://www.bioscience.heacademy.ac.uk/journal/vol1/beej-1-2.pdf].
121. Yaqinuddin A, Zafar M, Ikram MF, Ganguly P. What is an objective structured practical examination in anatomy? Anat Sci Educ, 2013, 6:125–133.
122. Yudkowsky R. Should we use standardised patients instead of real patients for highstakes exams in psychiatry? Acad Psychiatry 2002;26:187–92.
123. Zubair A, Chong YS and Khoo HE. Practical Guide to Medical Student Assessment. Published by World Scientific Publishing Co. Pte. Ltd. 5 Toh Tuck Link, Singapore 596224,

The World Journal of Medical Education & Research (WJMER) is the online publication of the Doctors Academy Group of Educational Establishments. It aims to promote academia and research amongst all members of the multi-disciplinary healthcare team including doctors, dentists, scientists, and students of these specialties from all parts of the world. The journal intends to encourage the healthy transfer of knowledge, opinions and expertise between those who have the benefit of cutting-edge technology and those who need to innovate within their resource constraints. It is our hope that this interaction will help develop medical knowledge & enhance the possibility of providing optimal clinical care in different settings all over the world.



WJMER

World Journal of Medical Education and Research

An Official Publication of the Education and Research Division of Doctors Academy

