

Framework for an effective assessment: From rocky roads to silk route

Zarrin Seema Siddiqui

ABSTRACT

A defensible assessment is the cornerstone of health professions education and requires a careful approach from all the stakeholders involved in designing and delivering the curriculum. In this article, a framework is described using four essential attributes that define an effective assessment. These attributes include orientation and clarification of expectations for both faculty and students followed by a well-structured learning experience and constant evaluation of the assessment program.

KEY WORDS: Assessment, Health professions, Medical Education.

doi: <https://doi.org/10.12669/pjms.332.12334>

How to cite this:

Siddiqui ZS. Framework for an effective assessment: From rocky roads to silk route. *Pak J Med Sci.* 2017;33(2):505-509.
doi: <https://doi.org/10.12669/pjms.332.12334>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

BACKGROUND

Effective assessment doesn't just happen. It emerges over time as an outcome of thoughtful planning...¹

Assessment is the cornerstone of educational cycle because it is the design of an assessment system that influences learning,² and therefore has been a constant focus of educational research. As far as health professionals are concerned, the educators are not only required to certify that a person should be awarded a qualification according to institution's legislations but they also have got responsibility to ensure that the certified health professionals are fit to deliver appropriate healthcare to the community. This implies that the assessors may need to defend

their decisions if a legal challenge occurs or an accreditation committee questions the assessment process. There is evidence that students identified as borderline of pass/fail decisions may graduate and cause concern regarding their competence.³ In another study, over thirty percent of assessors identified their reluctance to fail a student.⁴ Literature therefore recommends that assessment needs to be effective and that it is informed and governed by a set of principles which include alignment with the learning outcomes, fitness for the purpose of assessment, an assessment that is reliable, valid constant, fair transparent, inclusive and sustainable.⁵

In this article a framework for an effective assessment is proposed (Fig.1) with four essential attributes which include communication about the assessment, orientation to the assessment, embedding assessment within the learning experience and evaluation of the assessment. Each of these attributes is discussed in detail using examples from the author's personal experience and the literature.

Communication:

A clear communication between academics and trainees regarding the assessment requirements is the key to an effective assessment. For students it is important that they know what the assessment requirements are, what are the deadlines and what

1. Dr. Zarrin Seema Siddiqui, PhD.
Education Centre,
The University of Western Australia,
Perth, Australia.

Correspondence:

Dr. Zarrin Seema Siddiqui, PhD.
Associate Professor,
Education Centre,
The University of Western Australia,
M515, 35 Stirling Highway,
Crawley, WA 6009, Australia.
E-mail: zarrin.siddiqui@uwa.edu.au

* Received for Publication: January 11, 2017
* Accepted for Publication: February 28, 2017



Fig.1: COLE framework for an effective assessment.

is the criteria used. If there is any scaling of marks required or standard setting mechanisms applied, the details should be documented and available to students, staff and external reviewers.

One good example is the use of progress testing where students are assessed against the outcomes that are expected at graduate level.⁶ Students need to know why it is required in order to minimise the anxiety associated with assessment. The same holds true for the use of portfolios where medical students found the independent learning required to construct a portfolio as threatening given there was so much focus on knowledge acquisition.⁷ The assessment policy of the faculty should also be made available to students and staff so the requirements to progress within the course are known to all.

Orientation:

Besides communicating the expectations and requirements for assessment, orientation to a particular assessment is another crucial factor and this applies to both staff and students. There has been evidence in the literature that staff in clinical setting did not understand the assessment forms created by the educators as the terminology was not familiar to the clinical supervisors.⁸ They may interpret items on a form differently and may use diverse criteria for assigned ratings.⁹ Assessors must be fully conversant and trained with assessment methods used by the Faculty. For example, Vassiliou et al. introduced a new tool for evaluation of intraoperative laparoscopic skills.¹⁰ The authors ensured that all of the assessors

observing the performance of trainees have had experience with laparoscopic surgery and were familiar with the steps involved in the procedure. The observers were further trained by providing detailed instructions and reviewing the items on the checklist which resulted in an excellent internal consistency in terms of reliability. One study investigating perceptions of mentors regarding use of portfolios has suggested that if portfolios are used to promote self-directed learning skills among students, their mentors should be introduced to the relevant literature and develop their own self-directed learning skills. They have further added that mentors should experience the portfolio process themselves as learners.¹¹ Lennie and Juwah have recommended that faculty development in assessment be offered in areas such as pedagogy, range of activities, standardization, moderation of assessment activities, grading and feedback/feed forward after assessment.

Students also require orientation towards new assessment tools. For example, a student may have never experienced an Objective Structured Clinical Examination (OSCE) or oral examinations in their academic setting. This can provoke anxiety among students and affect the outcome of the assessment. Faculty should make provision in time table to acquaint students with assessment tools and grading criteria early in the course to minimize any stress. Use of formative assessment using the same tools will not only provide student with feedback but will also let them experience the assessment tool itself.

Learning Experience:

Assessment is now seen in terms of educational effectiveness and itself as a learning experience which provides the learners with tasks that are authentic, simulating what they are expected to perform in real life. This has led to introduction of formative assessments commonly known as assessment for learning, workplace based assessments where trainee has opportunity to be directly observed and receive feedback on his performance so he/she can improve on subsequent assessments.¹²

With the advent of technology the e-assessments and simulation based assessments are gaining popularity which is helpful to student as they provide instant feedback and have empowered students to be more responsible for their learning. There has been introduction of tools like portfolio based assessment where a holistic picture of the trainees is assessed for fitness to practice. Some

institutions are also offering Capstone experiences providing students opportunities to look beyond the medical model of practice.¹³

Evaluation:

Like all educational interventions and programs, evaluation of assessment is also required to assess if it is serving the purpose and is defensible. The characteristics that one looks for in the evaluation of an assessment are best defined in terms of the Utility formula and include estimation of reliability, validity, practicality, standardisation and educational effectiveness using quantitative and qualitative measures.¹⁴

Reliability:

Reliability of a test is a measure to produce similar results when applied at two different points in time.¹⁵ Let's say Huma administers a test to her students in Biochemistry. If Huma administers the same test to her students a second time, one would expect consistency in the scores received in two tests for each student.

Assessing a test's reliability is typically easier than assessing its validity. Statistical methods are used to measure the consistency within test results such as Alpha Coefficient or Kuder Richardson 21. Recently generalizability theory and D-study is also used for example Wilkinson et al used the same to report that for adequate reliability of DOPS, a trainee needs to be observed by a minimum of three assessors each observing a minimum of two procedures.¹⁶ Alves de Lima found that to achieve a reliable assessment using Mini CEX an approximate sample of nine encounters was required each with a different assessor, four encounters with two observers per encounter and three encounters when three observers are used.¹⁷

In performance based assessments, Objective Structured Clinical Examination (OSCE) has gained importance because of its reliability and the reason behind it is the predetermined criteria for each station, the sampling of clinical skills and cases that portray a more reliable assessment of a trainee and multiple assessors who are independently assessing at different stations minimising bias.¹⁸

Simple measures such as correlations between two scores given by a scorer at different times or correlations between two scorers can also be used to measure the reliability of an assessment. For example Ahsan makes a video recording of each of his dental students performing a high fidelity simulation. He grades each of the videos and then a month later, he examines the video recordings

again and re-scores them. He then computes a correlation between two sets of grades to see the correlation between his scores on for each student on two separate occasions. This is an example of intra-rater reliability.

On the other hand another example from a recent study is where authors developed a history and physical assessment form and used two paediatrician evaluators to assess ten student write-ups. They observed that the inter-rater reliability was higher for each item on the form compared to the overall assessment score.¹⁹

Validity:

Validity requires evidence from different sources that the assessment is measuring its intended purpose.²⁰ An assessment is not valid if it is assessing the professionalism using a written assessment because professionalism involves attributes that need to be observed in practice by multiple observers using multiple observations. Therefore structured or open ended questionnaires are used by members of healthcare team to inform 360 degree appraisal for specialist registrars.²¹ Likewise if standardised patients are exclusively used to assess the clinical competence of a trainee, it will not be a valid assessment of competence as the trainee will not receive timely feedback from the faculty as a result of direct observation of his/her practice.²²

There are different types of validity which are described as under.

- a. **Face validity:** an assessment is considered to have face validity if a number of judges ranging from experts to trainees agree that it is measuring what it intends to measure.
- b. **Content validity:** The content validity of an assessment takes into consideration both the curricular content and the expectations of the learners at various stages of their training in order to ensure that there is congruency in the learning outcomes and assessments offered to the students.²³ The process is often referred to blueprinting where faculty maps the content of the assessment against the learning outcomes using a matrix. External examiners can also retrospectively comment on the content validity of an assessment.
- c. **Construct validity:** Construct validity refers that the assessment is measuring the construct that it is intended to measure. An analysis of in-training examination for residents in radiation oncology demonstrated its construct validity as the

performance of residents improved with each year of additional oncology training but it was also observed that this effect persists only until level three of biology and physics training.²⁴

- d. **Concurrent validity:** is the extent to which an assessment correlates with other test of the same construct.²⁵ In a recent review, Mini Clinical Exercises used in the training of medical graduates have been found to have concurrent validity with the other tools measuring the clinical performance.²⁶
- e. **Predictive validity** is determined to assess if the said assessment can predict the future performance of the students. In a study measuring clinical performance by successfully completed clinical procedures in dentistry, it was observed that this measure has moderate positive correlation with performance in an OSCE.²⁷ Another study from UK demonstrated that the selection process used for General Practice training predict their future performance on MRCGP examination.²⁸
- f. **Practicality/Feasibility** A test which is highly valid yet too time consuming to administer or score, costly to purchase and is not usable is less likely to be acceptable by stakeholders. Among workplace based assessments, Direct Observation of Procedural Skills (DOPS) is considered a feasible examination as it requires only one assessor,²⁹ yet there is another study which considered feasibility as a problem as DOPS forms were completed and returned by only 33% of trainees in their study.³⁰ The same problem may also occur when not all the relevant procedures are encountered by a trainee in practice so they cannot be observed directly by an assessor.

Another example is the entry to postgraduate examination at the College of Physicians & Surgeons Pakistan in its formative years. The assessment required a written examination followed by a viva with four assessors. In the initial phase it was manageable but with the increasing number of applicants it became logistically difficult to manage and the assessment was changed to two written assessments comprising knowledge of basic sciences held at same day at various centres throughout the country.

One other illustration is of a medical course where Objective Structured Clinical Examination (OSCE) was held four times during the six year program. While it was considered a reliable exam, it was too time consuming and costly in terms of both physical and human resources hence the curriculum com-

mittee decided to reduce the number of OSCEs to two, with introduction of another assessment tool and more robust measures undertaken towards faculty training in development and administration so the assessment still maintains appropriate reliability and validity. One may argue that if OSCE is so costly then why not get rid of it altogether? The right answer will be that it is the cost effectiveness that compels faculty to persist with an OSCE because it clearly enables them to discriminate between different levels of performance of trainees.

At an individual level, consider example of Sonia who wants to measure clinical reasoning of her trainees. She considers developing a test where each student would select a case from their logbook to discuss in detail with Sonia in a one on one interview. Sonia feels that this test will have a higher content validity as she can assess the depth of knowledge. However as it will be too time consuming for her to administer she decided against it. Later she found that there is a commercial organisation that offers a test bank to be used which can assess the same however it would cost about 10\$ per student which is again not practical. Finally, Sonia developed her own written examination which comprises case based scenarios where students have to justify their responses as well as four structured short answer questions which was a more practical approach.

Therefore when an assessment is evaluated, the feasibility in terms of cost and acceptability is also considered.

Educational Effectiveness:

Feedback from students and faculty is also required to assess the educational effectiveness/impact of an assessment. There is evidence that students rely on what is covered in the tests/assessment for the purpose of what they need to learn. Usually these are part of the overall course evaluations. These should be embedded within the quality assurance processes looking explicitly at the assessment and be considered in making decisions. At the author's institution student representation is ensured in all committees looking at various aspects of curriculum so their feedback is continuously available and is acted upon as and when required.

CONCLUSION

An effective assessment is crucial to health professions education as this discipline has got the obligation to provide community with health professionals who are safe and competent practitioners. A framework has been presented which includes all the attributes that contribute to

an effective assessment, i.e. the expectations are communicated in a timely way with appropriate orientation to both students and staff and is embedded within learning experience with opportunities for feedback with constant evaluation of the process and product. Yet it should be kept in mind that in real world an educator is always faced with problem as to how they maintain the adequate balance between different constructs and a compromise is inevitable depending on the purpose of assessment and the regulations of the institutions.

Grant Support & Financial Disclosures: None.

REFERENCES

- Banta T, Jones E, Black K. Designing Effective Assessment. San Francisco: Jossey-Bass; 2009.
- Kaufman D. ABC of learning and teaching in medicine: applying educational theory in practice. *Br Med J*. 2003;326:213-216.
- Tamblyn R. Association between licensing scores. *JAMA*. 1998;280:989-996. doi:10.1001/jama.288.23.3019.
- Morton J, Cumming A, Cameron H. Performance-based assessment in undergraduate medical education. *Clin Teach*. 2007;4(1):36-41. doi:10.1111/j.1743-498X.2007.00138.x
- Lennie S, Juwah C. Exploring assessment for learning during dietetic practice placements. *J Human Nutr Dietetics*. 2010;23(3):217-223. doi: 10.1111/j.1365-277X.2010.01075.x.
- Schuwirth LWT, van der Vleuten CPM. The use of progress testing. *Perspect Med Educ*. 2012;1(1):24-30. doi:10.1007/s40037-012-0007-2.
- Snadden D, Thomas M. The use of portfolio learning in medical education. *Med Teach*. 1998;20:192-199. doi:10.1080/01421599880904.
- Rutkowski K. Failure to fail: assessing nursing students' competence during practice placements. *Nurs Stand*. 2007;22(13):35-40. doi:10.7748/ns2007.12.22.13.35.c6299.
- Fadheela A-M. Construct validity and generalizability of Pediatrics clerkship evaluation at a problem-based medical school, Bahrain. *Eval Health Prof*. 2009;32(2):165-183. doi:10.1177/0163278709333149.
- Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. 2005;190(1):107-113. doi: 10.1016/j.amjsurg.2005.04.004.
- van Schaik S, Plant J, O'Sullivan P. Promoting self-directed learning through portfolios in undergraduate medical education: The mentors' perspective. *Med Teach*. 2013;35(2):139-144. doi:10.3109/0142159X.2012.733832.
- Gavriel J. Assessment for learning: a wider (classroom-researched) perspective is important for formative assessment and self-directed learning in general practice. *Educ Prim Care*. 2013;24(2):93-96. doi:10.1080/14739879.2013.11493462.
- Tewfik D. Collaboration in the community: A Capstone experience. *Occup Ther Health Care*. 2013;16:111-117. doi:10.1080/J003v16n02_10.
- Van der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ Theory Pract*. 1996;1(1):41-67. doi:10.1007/BF00596229.
- Reznik R. Teaching and testing technical skills. *Am J Surg*. 1993;165:358-361. doi:10.1001/archsurg.1997.01430260098021.
- Wilkinson J, Crossley J, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ*. 2008;42:364-373. doi: 10.1111/j.1365-2923.2008.03010.x.
- Alves de Lima A, Conde D, Costabel J, Corso J, Vleuten C. A laboratory study on the reliability estimations of the mini-CEX. *Adv Health Sci Educ*. 2013;18(1):5-13. doi:10.1007/s10459-011-9343-y.
- Boursicot K, Etheridge L, Setna Z, Sturrock A, Ker J, Smees S, et al. Performance in assessment: Consensus statement and recommendations from the Ottawa conference. *Med Teach*. 2011;33(5):370-383. doi:10.3109/0142159X.2011.565831.
- Middleman AB, Sunder PK, Good Yen A. Reliability of the history and physical assessment (HAPA) form. *Clin Teach*. 2011;8(3):192-195. doi: 10.1111/j.1743-498X.2011.00459.x.
- Schuwirth LW. Assessing medical competence: finding the right answers. *Clin Teach*. 2004;1(1):15-18. doi: 10.1111/j.1743-498X.2004.00012.x.
- Carr S. Assessing clinical competency in medical senior house officers: how and why should we do it? *Postgrad Med J*. 2004;80:63-66. doi:10.1136/pmj.2003.011718.
- Holmboe ES, Ward DS, Reznick RK, Katsuftrakis PJ, Leslie KM, Patel VL, et al. Faculty Development in Assessment: The Missing Link in Competency-Based Medical Education. *Acad Med*. 2011;86(4):460-467. doi: 10.1097/ACM.0b013e31820cb2a7.
- McLaughlin K, Lemarie J, Coderre S. Creating a reliable and valid blueprint for the internal medicine clerkship evaluation. *Med Teach*. 2005;27(6):544-547. doi:10.1080/01421590500136113.
- Paulino AC, Kurtz E. American College of Radiology In-training examination for residents in radiation oncology (2004-2007). *Int J Radiat Oncol Biol Phys*. 2008;70(3):666-670. doi: 10.1016/j.ijrobp.2007.09.049.
- Batchelder AJ, McCarthy MJ. A Critical Appraisal of the Evidence of Current Methods of Assessment for Surgical Trainees. *Bull Royal Coll Surg Engl*. 2013;95(6):7-11. doi: 10.1308/003588413X13643054410304.
- Ansari AA, Ali SK, Donnon T. The Construct and Criterion Validity of the Mini-CEX: A Meta-Analysis of the Published Research. *Acad Med*. 2013;88(3):468-474. doi:10.1097/ACM.0b013e318280a953.
- Graham R, Zubiaurre Bitzer LA, Anderson OR. Reliability and predictive validity of a comprehensive preclinical OSCE in dental education. *J Dent Educ*. 2013;77(2):161-7.
- Ahmed H, Rhydderch M, Matthews P. Do general practice selection scores predict success at MRCP? An exploratory study. *Educ Prim Care*. 2012;23(2):95-100. doi:10.1080/14739879.2012.11494083.
- Hamilton K, Coates V, Kelly B, Boore J, Cundell J, Gracey J, et al. Performance assessment in health care providers: a critical review of evidence and current practice. *J Nurs Manag*. 2007;15:773-791. doi: 10.1111/j.1365-2934.2007.00780.x.
- Wilkinson JR, Crossley JGM, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ*. 2008;42(4):364-373. doi: 10.1111/j.1365-2923.2008.03010.x.