

Assessment and evaluation in medical education

Gibbs T, MB Ch.B (Liverpool), FRCGP, D.A (Educational Leadership), M.Med.Sci, ILTM, F.F.Hom
Professor of Medical Education and Family Medicine, College of Medicine and Medical Sciences,
Arabian Gulf University, Kingdom of Bahrain

Brigden D, MEd, MI Biol, ILTM
Visiting Professor of Professional Development, Faculty of Health Sciences,
University of Cape Town & Adviser for Postgraduate Medical and Dental Education,
Mersey Deanery, University of Liverpool Liverpool, UK

Hellenberg D, MB Ch.B (UCT), M.Fam.Med, MFGP (SA)
Professor of Family Medicine, University of Cape Town, Faculty of Health Sciences, University of Cape Town

Correspondence: Professor Derek Hellenberg, Professor of Family Medicine, University of Cape Town,
Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, 7925, Cape Town.
Tel: 021 406 6712 FaX: 021 406 6244 e-mail: dhellenb@cormack.uct.ac.za

Summary

Previous discussion has focused upon the educational process, specifically, the basis for “good” medical education, the methods of learning and teaching used and the relationship between undergraduate, postgraduate and life-long learning; continuing professional development. All these would appear to be, in the first instance, linear processes, with a beginning that commences with a specific teaching and learning activity and an end, often an end-point examination, hopefully testing how much we remember, apply and use.

As part of any educational programme, assessment and evaluation feature frequently. The purpose of this article is to not only equip the reader with an understanding of the terminology used in assessment and evaluation, but to demonstrate that, as for many other aspects of medical education, the processes of assessment and evaluation bring cyclical and dynamic actions which can be used and built upon to improve the efficiency of the training model and even oneself.

“Examinations, sir, are pure humbug from beginning to end. If a man is a gentleman, he knows quite enough, and if he is not a gentleman, whatever he knows is bad for him. “

Oscar Wilde¹

(SA Fam Pract 2006;48(1): 5-7)

Defining the Terms

Medical education is similar to many other branches of health care, enriched with words that its participants often use interchangeably and without fully understanding their meaning. Like health care, where clarity of understanding produces high quality service, a firm understanding of terminology is absolutely necessary in medical education.

It is generally accepted that, despite many years of discussion, assessment and evaluation remain poorly understood topics and are frequently undefined.

The philosophical debate moves around whether assessment is *of* learning, a measurement of outcomes, or *for* learning, assessment driving the curriculum. Similar debate focuses on evaluation, whether it is really just another word for assessment or is used to give real value to any

educational activity.

Assessment, derived from a Latin word that means “to sit beside and judge” is appropriately used to describe the systematic gathering of information about what the learner should know, able to do or working towards. It is usually associated with some measurement, marks or percentages, but could be associated with specific descriptors, excellent, good, average or poor. When describing assessment, it is difficult to avoid the concomitant use of other terms. If assessment is a measure we need to know what it measuring. *Competency* is one such measure and refers to the specific skill that has been taught, and is now being measured. *Competency* and *objectives* or *outcomes* are often used interchangeably in describing what has to be measured. This form of measurement has to be judged

against a certain *standard* or *benchmark* to be achieved. In describing the students’ achievement, it is common to refer to these standards as well as describing its reference point. If we measure and compare student performance against a large number of his peers and judge how or where he fits into the distribution of marks, this is known as *norm referencing*. This is used if we wish only a certain percentage of students to pass or fail. However, if we assess how students perform against a specific objective or criterion, this is known as *criterion referencing*. With this latter form, all students can pass an assessment if they have all reached the required *standard* or *competency*.

When talking of assessment, three forms are commonly used. *Formative assessment*, whereby assessment is often informal, continuous or on-going and is a two- way communicative

process between teacher and learner/s, with a large degree of discussion regarding strengths, weaknesses and opportunities to improve. *Summative* or *end-point assessment* is much more formal, and judgmental, (the traditional examination system). It tends to be more numeric and *quantitative*, whereas formative assessment tends to be more descriptive and *qualitative*. *Self* or *ipsative assessment* concerns measuring oneself; you learn about yourself, you test yourself against standards or specific parameters, giving you an idea of how you are performing. This form of assessment becomes very important in teaching methods such as problem-based learning, where *self-directed learning* often leads to students finding difficulty in knowing or understanding to what level of learning they need to work towards.

It is interesting to see how, in recent years, other forms of assessment are being developed, and readers are directed to the suggested reading section, specifically related to *authentic* or *real-world assessment*. Here, students are assessed on their performance in realistic or real-life settings, often known as *high-fidelity simulation*.

Finally, when assessment is being discussed, we have to consider whether the tools used to assess are reliable and / or valid? *Reliability* is the extent to which a test of assessment is repeatable, reproducible and yields consistent scores. All measurements have the potential for error and scores can be variable, based upon this error. Errors are often random, poor lighting, noisy exam room, student anxiety, illness etc. *Reliability* is a measure of how much the score is influenced by these error potentials; hence a reliable test has a low error score. The *validity* of an assessment is the extent to which a test really measures what it is supposed to measure. It is important to note that a test can be reliable, but not always valid.

If the assessment looks right and looks to be measuring the right thing, it is said to have *face validity*. If the

content is appropriate and actually measures what it purports to measure, it is said to have *construct validity*. If an assessment test of a specific competency has the ability to predict performance in other areas or competencies, it is said to have *criterion validity*.

A brief word on *generalisability*. Often we want to know whether the results of a measure or test used with a particular group can be applied to other tests or measures. This is known as generalisability. If we have two measures of IQ, we would wish to know if both measures give the same result. If they do, they are said to be *generalisable*.

Evaluation is a much broader concept, measuring the value of educational activities, programmes, curricula etc. Whereby assessment relates to individuals or groups and their performance, evaluation relates more to process, content and appropriateness of activities. A typical example of evaluation in action is the enquiry made of a specific course:

- did it have the required effect?
- did it address the correct issues?
- were the needs and the wants of the teachers and learners achieved?
- will it require changes if performed again?

One can see from these examples that true evaluation of programmes/ educational activities requires enquiry of both teacher and learner and results in potential for change in consequent

activities: it is a *cyclical and dynamic process*. A *by-proxy* measurement of how well an educational activity has achieved its expected outcomes could be obtained by judging how well students have performed; if students perform well, the course has achieved its aims. This measurement is weakened by the failure to specifically ask the participants (teacher or learner) how they felt about the educational activity! Like summative assessment, evaluation is usually conducted at the end of a specific educational activity, however to be efficient, evaluation is intrinsic to each separate step or component of that activity. Evaluation requires *clarity* (all can see the results, the results are public property), *credibility* (something happens or changes if the evaluation suggests it should) and *consistency* (it becomes an important feature of all educational activities).

Consequent upon the above descriptors, no educational activity is complete without an assessment and evaluation procedure, each achieving its own clarity of action, credibility of reaction and consistency of findings.

The characteristics of a successful assessment and evaluation system are found in figure 1, and it follows that:

- Assessments should focus upon performance in as near real life activity as possible.
- The assessment task should be rooted in the real-life or clinical practice of both the assessed and the assessor
- A variety of sources and methods

Figure 1: Characteristics of a successful assessment and evaluation system

- It is based upon objectives that are clearly understood and agreed by the assessor and the assessed
- Uses methods that match closely the objectives of the educational activity.
- Provides a measure of performance and effectiveness at the onset and end of training
- Is manageable within the constraints of the working environment
- Is simple and robust enough to cope within varying educational contexts
- Is easy to understand
- Is used consistently
- Is based on broad, but explicit, criteria
- Is perceived by its users and others as a valid approach
- Distinguishes, as far as possible, developmental and judgmental processes
- Enhances learning in all participants
- Provides evaluatory feedback to all participants.

- of assessment should be used
- The tasks should be manageable in the context of service or clinical provision
- The assessment tasks should not consume excessive resources
- A system of regular enquiry should be used to ensure the procedures are valid and reliable
- Each assessment that completes an educational activity is scrutinised by a purposeful and active evaluation.

Assessment and evaluation in everyday practice

This article has tried to introduce the reader to some of the basic ideas, definitions and applications of assessment and evaluation. The reader is directed to further reading should they require a more in-depth knowledge of the subjects.

The vast majority of the readers will have encountered the concepts sometime during their undergraduate and postgraduate experiences, either as a teacher or learner; but can we apply the defined principles in everyday practice? Should we be constantly assessing ourselves and our peers? Should we be reviewing or evaluating our daily activities, personal, practical or managerial? Can the principles of assessment and evaluation form the basis of appropriate continuing professional development? If accreditation procedures become accepted and commonplace, will peer and self-assessment be intrinsic to the process?

Perhaps we should not consider assessment and evaluation as only relevant to educational activities and divorce them from practice; perhaps they are intrinsically and appropriately connected to our daily lives. 🙌

“Examinations are formidable even to the best prepared, for the greatest fool may ask more than the wisest man can answer.”

Charles Caleb Colton²

Points to Ponder

- Assessment and evaluation are key concepts and components of any educational activity.
- Clarity of definition and application aid understanding of the concepts.
- Both concepts may be applied to daily practice to enhance personal development.

References

1. Oscar Wilde (1856 - 1900) Anglo-Irish playwright, novelist from "The Picture of Dorian Gray," 1891.
2. (1780-1832), English, author, philosopher and eccentric

Suggested Reading

- a. David T, Patel L, Burdett K, Rangachari P. Problem-Based Learning in Medicine. The Royal Society of Medicine Press. UK. 2005.
- b. Freeman R, Lewis R. Planning and Implementing Assessment, London, Kogan Page. London. UK. 1998
- c. Gronlund NE, Linn RL. Measurement and Evaluation in Teaching. Macmillan Publishing Company. New York. 1990.
- d. Linn RL, Gronlund NE, Davis KM. Measurement and Assessment in Teaching. Prentice Hall, New York. 1999
- e. Peyton JWR. Teaching and Learning in Medical Practice. Manticore Europe Limited. UK 1998