

Issues in assessment and appraisal in internship training

I.G. Premadasa

Examiners should primarily be concerned with the validity of the test when selecting instruments of assessment. Assigning a numerical value does not increase the precision of the measurement, and they should be wary of making fine distinctions among candidates based on such results. Formative assessment gives the opportunity for the trainee to identify the content areas in which achievement is poor, for possible remedying, while summative assessment allows for summing up the total amount

learned. Clinical skills are not transferable between patients, performances being case specific. Sampling the content widely increases the accuracy of the evaluation decisions.

Key words: Internship, assessment, appraisal, validity, reliability

Bull Kuwait Inst Med Spec 2003;2:19-21

Introduction

The internship year constitutes an important phase in the transformation of the student doctor in his undergraduate days to the fully-fledged medical practitioner. Internship training is one of the main areas that comes within the mission of Kuwait Institute for Medical Specialization (KIMS).

In its efforts to upgrade the internship training program, KIMS has taken many steps recently. Under curriculum review, the outcome objectives that the trainees are expected to acquire in each specialty have been listed, and a format has been designed for use as a record of formative and summative evaluations.

Established tradition exerts an overriding influence on many of the practices and procedures that are used today in teaching and assessment of clinical competencies. Though approaches such as the OSCE (objective structured clinical examination) have received wide acceptance in medical academia, there is a need for greater insights into the information available in medical education research and its incorporation into curriculum planning and application in teaching and assessment.

Some issues that surface when discussing assessment, in general as well as of clinical competencies, are analyzed below. Also listed are a few topic areas that need to be included in the training program, especially when planning assessment.

Assessment

Assessment involves determining the presence of a previously listed ability in a learner, and quantifying it. The measurements need to be *objective* (different examiners agreeing on the score or grade) and *reliable* or reproducible (similar readings obtained on repeated measurement). Usually the results are judged against some criteria, in which case the process would rightly be called evaluation. When assessment is carried out at the end of course unit or a program, it enables summing up of what has been learned, and the information is used for certification. Assigning a numerical value to identify a grade does not increase the precision of the assessment, and trainers and examiners should avoid making fine distinctions among numbers derived in this manner.

Though it may be possible to obtain assessments that are highly objective and reliable (such as by using MCQs), the most important consideration in assessment is whether the assessment tool is being used to measure what it has been designed to measure. In other words, is the assessment *valid*? This means that when we assign a score or a grade

Director, CME Center, Kuwait Institute for Medical Specialization.

Correspondence: Dr. I.G. Premadasa, Kuwait Institute for Medical Specialization, P.O. Box 1793, Kuwait 13018. Tel: 965 2410027 Ext. 107; Fax: 965 2467140; Email: cmecenter@kims.org.kw

to a trainee, we should be certain, as far as practically feasible, that those who perform well on the assessment have acquired the ability to a satisfactory degree. For instance, if a trainee's oral communication skills with the patient are to be tested, then he or she should be placed in a situation in which the examiner could observe the candidate communicating. Giving a written test, where the candidate describes how he would communicate is an invalid form of testing.

Reliability is a theoretical concept that assumes that the amount that has been learned does not change between the two (or more) occasions. Additionally, examinations cannot be repeated merely to satisfy the dictates of test specialists, but thankfully, mathematical approaches that 'simulate' test-retest situations are available.

Reliability is concerned with whether the results are reproducible when the examiner assesses the candidate again, while objectivity refers to reproducibility when different examiners use the test. The mere reproducibility or repeatability of test results should not be taken as indicating that the test is 'good', or that it is being used correctly: two examiners at an oral examination may assign similar marks to an examinee biased by prior knowledge or unduly influenced by the language skills of the candidate.

Repeatability also does not always guarantee accuracy. When experienced examiners were asked to measure the length of a table using a tape measure from which the first inch had been removed without their knowledge, they invariably came out with the same reading—an objective, but inaccurate result

Appraisal

Appraisal usually involves a description of activities, achievements and characteristics that have resulted from an analysis involving the subject of the appraisal. Its main purpose is to record achievements and chart the progress of the trainee. It also allows the trainee to agree on future goals that he or she may aspire to. Appraisal therefore allows the trainee to receive comprehensive and timely feedback, so that steps may be taken to deal with any deficiencies in achievement.

Two terms that need clarification are formative and summative assessment. Forma-

tive assessment is purely for feedback, so that the trainee could improve in weak areas, before it is too late. Therefore, formative assessment has to be undertaken during the training program. Summative assessment, on the other hand, is for certifying the level attained, and is done on completion of training. The results may be indicated by a pass or fail label, or as a point in a grading scale. The two assessments perform opposing purposes, and therefore, the same assessment cannot be used for serving both ends – in formative, the candidate benefits most if he exposes all the weaknesses, while in summative, he would receive a high score by revealing all the strengths and hiding the weak areas.

Current Assessment Procedures

An important aspect of clinical performance that has been widely researched over the past two decades is content specificity. The results of these investigations have shown that the ability to perform well in one clinical problem does not transfer to others.^{1,2} It means that a trainee who scores high when given one clinical problem may not do so when faced with another clinical topic or when required to perform another procedure. This observation has seriously questioned the wisdom of relying on one 'long case' at a summative assessment, prompting examiners to use a larger number of 'cases' and widening the sampling.

Objective tests such as MCQs are far removed from the reality of the clinical setting. Some researchers have even gone to the extent of saying that 'patients don't present with five choices' and have offered alternative strategies.³ Additionally, an excessively large proportion of MCQs, as they are used at present, assess the trainees at only the level of recall of data. This is mainly because writing items to test at higher levels is not a competency that many trainers and examiners have mastered. It may even be seen that the questions referred to as 'problem-solving', popular with trainers in clinical disciplines, are often not written at the level problem-solving in the cognitive domain.

Approaches to Assessment and Appraisal

The widely used format of the final examination in the clinical disciplines cannot be rated

high on measurement characteristics when one considers its low fidelity and inadequacy of sampling, and analyzes it from the point of view of content specificity. Evidence on the trainee's progress and level of achievement obtained from a variety of sources is therefore needed to make an informed judgment. Records maintained by the trainee in the form of log books and portfolios, coupled with information provided by those not directly connected with the collection of assessment data would be useful in this activity. Multi-station examinations and ward ratings accumulated over the duration of the attachment are other useful sources of information.

Areas to be Included in Internship Training

In addition to the medical knowledge and purely clinical skills, a number of other areas need to be included in the assessments. When determining what has to be assessed it is important to keep an appropriate balance between the requirements of the new SHO or Registrar and the life-long learning needs of the practitioner.

Communication skills, values and virtues, attitudes and professional behavior, computer literacy, information retrieval, learning skills, reflective abilities and team working skills are examples of some content areas that should be emphasized during internship training. Others that need to receive due consideration are personal development and career orientation.

References

1. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination, *Med Educ* 1988;22:325-34.
2. Elstein AS. Beyond multiple choice questions and essays: the need for a new way to assess clinical competence, *Acad Med* 1993;68:244-9.
3. Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. Patients don't present with five choices: an alternative to multiple choice tests in assessing physicians' competence. *Acad Med* 1999;74:539-46.

Further reading

Lowry SR, Buckley EG, eds. *Conference on the Transition from the Medical Student to the Medical Practitioner: Proceedings*. Edinburgh: Association for the study of Medical Education; 1997.