Assessment of the general practitioner

H MULHOLLAND

P M J TOMBLESON

SUMMARY. There is considerable interest in general practice assessment both as a means of monitoring vocational training and as a form of self audit for the general practitioner. Basic principles involved in the selection and design of assessment procedures are reviewed and their application discussed.

'Without doubt, the assessment system is the most potent factor influencing student learning behaviour'.[1]

Assessment of the trainee practitioner

SUMMATIVE assessment engenders controversy in all fields of education; family medicine training programmes are no exception, with the trainees submitting themselves to a pass/fail assessment against their peers, and trainers feeling that the effectiveness of their teaching is under scrutiny. There are several well documented studies[2] confirming the desire of students to be assessed at the end of a period of training although an effective programme of formative assessments may negate this effect (Marshall J, Australian family practice training programme, personal communication). Indeed, a study of 280 trainees in 1985 showed over two thirds intending to take the MRCGP examination as 'a personal hurdle or discipline' while perception of the degree as 'helpful in getting a job' was an equally important factor, especially among women.[3]

The nature of assessment not only affects the knowledge and skills that students acquire but also their approach to learning.[4] Tests confined to factual recall (such as traditional multiple choice questions) tend to induce a superficial 'rote-learning' approach to the student's learning whereas those requiring a greater depth and range of understanding encourage a more thorough approach in which the learner tries to arrive at real comprehension. Quite reasonably, this influence may lead to perceptions of a 'back-wash effect' whereby the examination appears to dictate the curriculum;[5] examiners thus need to be sensitive to the content areas covered, particularly where no detailed teaching curriculum exists on which to base the assessment.

Framework for assessment

There have been many attempts to define the work of the general practitioner yet new formulations continue to be produced giving rise to speculation about the degree of comprehensiveness that can be achieved in any assessment process. Models from the field of education have been used as the basis for some developments. Bloom's taxonomy[6] of educational objectives defined cognitive, affective and psychomotor domains as the basis for learning, and therefore of assessment. Fabb and Marshall[7] referred to the interplay of these attributes with a further dimension of health management skills for different groups of patients and a third dimension of 'clinical competence' (Figure 1). It is this three dimensional model which has served as the basis of many developments in general practice education.

H Mulholland, MA, PhD, lecturer, Centre for Medical Education, University of Dundee. P M J Tombleson, MBE, FRCGP, convenor of the panel of examiners, Royal College of General Practitioners.
Submitted: 6 July 1989; accepted: 14 September 1989.

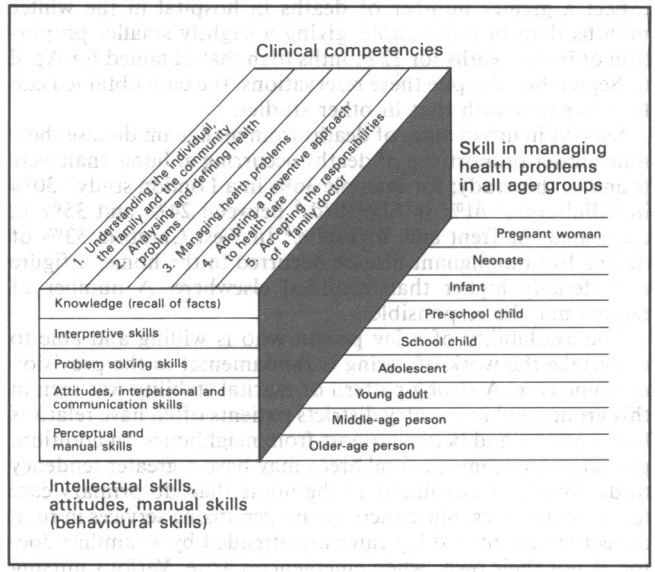

Figure 1. *Three dimensions of general family practice (reproduced with permission of the Royal Australian College of General Practitioners).*

Candidates might, for example, be questioned on their possible responses to a pregnant woman who requests home confinement. Their answer would incorporate their understanding of the management issues raised by home confinement and their clinical judgement. This three dimensional diagram allows us to crystallize and combine many previous models and lists of 'competence', 'attributes' or 'content areas' involved in general practice. We suggest that the three dimensions might be relabelled as content, tasks and personal characteristics (Figure 2).

Ideally, in practice the three circles should overlap so that the centrality of the patient assumes increased importance. Experienced family doctors function with the patient calling on their knowledge, ability to perform tasks and personal characteristics in a unified doctor–patient mode. Unfortunately, no examination can directly test this real world performance. In the case of the MRCGP examination the aspects of competence have therefore been separated and a variety of assessment instruments devised, each of which claims to test one or more of these aspects. Unfortunately, experienced principals sometimes do less well than expected in the MRCGP examination. They are used to working in the integrated mode and may find it difficult to separate out the different aspects of competence. Examples of the three aspects are:

- Content area — oncology, rheumatology, the cardiovascular system.
- Tasks — history taking, physical examination, writing a practice report.
- Personal characteristics — knowledge, dexterity, integrity.

Using such a model to define the areas being assessed will ensure that an adequate range of content, tasks and attributes are sampled and so provide a more structured framework to guide decision making.

Validity

The model shown in Figure 2 provides us with the capacity to answer the most important question in any assessment design

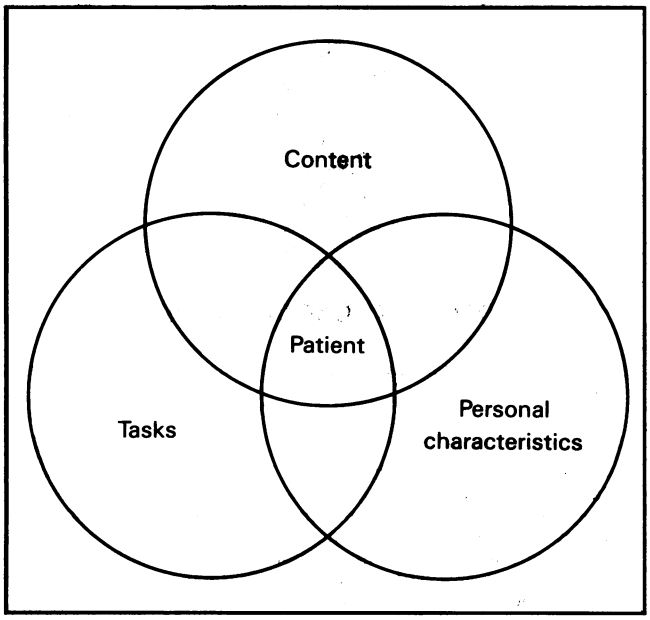

Figure 2. *The three aspects of competence.*

— 'what is being assessed?' The aspect of an assessment process which relates to this question is 'validity'; a test is valid if it tests what it is meant to test. Validity comes in many guises. The test has 'content validity' if it truly reflects the work of general practice. A test may have 'face validity' if it is perceived to test this area, but only careful analysis of the candidates' scores will show if they are actually reflecting real world performance. 'Construct validity' is of importance if what we want to assess is not clearly observable in the natural world. Attributes such as empathy cannot be measured with weights or yardsticks; definitions of 'empathy' will test the construct. However, since each of us has created our own construct based on experience, the tests we use to assess 'empathy' are likely to be varied and wide ranging. 'Predictive validity' is the desired attribute of any summative assessment. Do candidates who pass the examination actually go on to perform better than those who fail? Unfortunately research evidence in the area of predictive validity is difficult to obtain and has sometimes been collected on the basis of highly individual constructs of what constitutes effective real world performance.

For each test the model can be used to decide 'what content area is being assessed?', 'what tasks are being observed?' and 'what are the underlying attributes which will help the candidate to gain marks on this test?'

Failure to assess the full range of competences may be related to the temptation to use easily measured criteria. Thus, a general practice assessment may overweight evidence of immunization take-up rate, cervical smear coverage or good note-keeping because measurement of consultation skills or of the handling of sensitive ethical issues is too difficult.

Nature of assessment

Assessment may be either formative or summative. Formative assessment is designed to provide a basis for improving the learner's performance through feedback. In the MRCGP examination the assessment is summative, describing the attainment of the candidate at one point in time. Criterion referenced measurements compare the learner with a set of well defined descriptions of performance, whereas peer-referenced (or norm-referenced) measurements compare each learner with other members of the group. The assessment of vocational training should ideally be criterion referenced; it should say whether or not the trainee is able to achieve certain specific goals. Criterion referenced measures of clinical competence are difficult to design; the Board of Accident Surgeons examination in the USA is a rare example. Some examinations, including the MRCGP examination, have introduced criteria as mandatory pass/fail features, for example cardiopulmonary resuscitation.

Reliability

Any assessment should be reliable — the test should produce the same results if used with different markers on different occasions. The pursuit of reliability has caused the multiple choice question format to be adopted in some examinations, particularly in the USA, to the exclusion of other valid techniques. Because its reliability is high there can be no question of marker error or bias.

If a test is not reliable it cannot be valid. If a candidate scores 20% on one occasion and 50% on another with the same test, we do not know how the marks are being awarded and so cannot know what is being measured. Any summative assessment must be based on reliable measuring instruments.

Global assessment, where one person passes a single judgement of the candidate (as with trainer/trainee assessment using rating scales) is often found to be unreliable. One factor in this unreliability is the 'halo effect' whereby attributes valued by examiners may impress them to the exclusion of areas of deficiency, which they may possibly share. Problems may also occur in comparing ratings given by different trainers. Rating scales can be used in such a way that most trainees are rated as close to the average and this will hide the real strengths and weakness of individuals.

Feasibility

Any assessment method must be feasible. Can the practical requirements of the examination be met in terms of accommodation for candidates, examiners, support staff and if necessary, patients? Time constraints may dictate certain actions. One thousand multiple choice question papers may be marked in one hour by an optical scanner whereas direct observation of the candidate working in a surgery will demand at least a half day of observation by at least one examiner per candidate. Feasibility problems may preclude a thorough assessment in certain areas and thus manual skills, such as examination of the prostate or injection of a frozen shoulder, may be impossible to observe directly. It is for these difficult areas of assessment that new techniques need to be considered.

Generalizability

The effect of making judgements based on inadequate information may be compounded by our failure to recognize that doctors vary considerably in their skills between different areas of competence. This concept may come as a surprise to many examiners, yet it is obvious that there is more than one way to be a good general practitioner. In such a wide field there is a paramount need for breadth of questioning and a clear definition of the areas to be explored. By the use of analytical research techniques and complicated statistical procedures, such as generalizability theory, we may be able to develop a reliable and valid assessment of competence. Generalizability theory is based on the concept of finding out how many samples need to be taken from the universe of all possible general practice activities before a clear and reliable picture of the candidate's performance can be obtained.

Conclusion

The development of alternative assessment instruments, such as the objective structured clinical examination,[8] continues to shed light on hitherto unexplored aspects of the family doctor's range of skills. The need for a summative assessment, such as the MRCGP examination, at the end of vocational training should theoretically become less as reliable forms of formative assessment become part of our national training programme. Nevertheless, it has to be said that the excellent Australian model of formative assessment has been deemed insufficiently reliable to guarantee quality and a summative examination reinstated. In the meantime, work will continue to improve the MRCGP examination by introducing reliable methods of summative assessment in newly defined areas of general practice. By using the three axes of the model given here it is hoped to develop a method of defining these areas.

References

1. Elton LRB, Laurillard DM. *Learning style in continuing medical education.* Toronto: Council on Medical Education, Canadian Medical Association, 1979.
2. Scottish Education Department. *Assessment for all (Dunning report).* London: HMSO, 1977.
3. Tombleson PMJ, Wakeford R. Why do trainees take the MRCGP examination? *J R Coll Gen Pract* 1989; **39**: 168-171.
4. Marton F, Saltor R. On qualitative differences in learning. *Br J Educ Psychol* 1976; **46**: 115-127.
5. Department of Education and Science. *Glossary of terms prepared by the Joint Council of GCE and CSE Boards for their subject working parties for national criteria.* London: HMSO, 1981: 20.
6. Bloom BS. *Taxonomy of educational objectives.* New York: David Macray, 1983.
7. Fabb WE, Marshall JR. *The assessment of clinical competence in general family practice.* Lancaster: MTP, 1983.
8. Harden RM, Gleeson FA. Assessment of medical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979; **13**: 39.

Address for correspondence

Dr H Mulholland, Centre for Medical Education, Ninewells Hospital and Medical School, Dundee DD1 9SY.