Work-based competency assessment: the way forward for run-through specialties?

Work-based assessments have been introduced recently into the foundation programmes and may be extended further into hospital training. They have not been quality assured. The same standards should be applied to work-based assessments as those for national postgraduate assessments. All may be built into a competency framework.

The process for devising and setting assessments is clear. It occurs in three parts: the designing of the curriculum and its assessments, its delivery, and then the evaluation of the assessments. The evaluation should include feedback to the student. There is a long history of assessment predating the idea of competency, and confusingly various authorities use 'competency' differently. The most frequent variation in the work place is to use competency instead of skill. Authorities talk of a list of competencies at a level. Competency is best considered as a framework around which assessments are devised as this reflects its derivation. Work-based competency assessments in medicine must be fitted into this framework along with established frameworks.

The whole process of competency assessment is both complex and difficult, and is now bogged down in ideology, which is often politically driven. The quest to measure competency started with a search for the best way to evaluate the qualities an individual required for a job. David McClelland (1973), a behavioural psychologist and sociologist who worked at Harvard, developed the concept from his studies on educational measurement. He considered that the reliance on intelligence testing was inappropriate and that behavioural analysis based on criterion sampling should be used. His background meant that he attached great importance to outcomes, validity and reliability: difficult to achieve in work-based assessments. He also believed that we should move away from high-level skills which cannot be measured, such as knowledge, skills and attitudes, towards the characteristics that make up an occupation.

This latter methodology is the approach of psychometric testing, which relies on the measurement of knowledge, abilities, attitudes and personality traits, which looks at the differences between individuals and again was developed from intelligence testing. A recent article in the *British Medical Journal* commented that we require a number of assessments if we are to do this (Schuwirth et al, 2006).

Mr Adrian Drake-Lee *is Consultant Ear, Nose and Throat Surgeon, University Hospital NHS Trust, Birmingham B15 2TH*

Building a framework

It is difficult too to know what we are measuring when a complex domain is measured. The domain 'attitudes' is high level and comes from judgment. It may be thought of as containing an ABC just like resuscitation: affect, behavioural change and cognition. Cognition too can be broken down into memory, attention, perception, action, problem solving and mental imagery. All may be tested if we wish. On the other hand at the simplest level, every assessment is an assessment of a skill. We could use these three domains – knowledge, skills and attitudes – for our competency framework if we modify them and make them more explicit: looking for the characteristics of each specialty (*Figure 1*).

Knowledge may be replaced with medical knowledge. Skills could be replaced with practical or manual skills (a type of ability) and attitudes may be replaced by clinical behaviour. These three divisions reflect how clinical medicine is practiced. The combination of medical knowledge and clinical behaviour fits in with Miller's (1990) simple pyramidal model of competence where cognition comes before behaviour. Clinical behaviour may be broken down into required traits or areas such as history taking, communication skills, decision making and problem solving. These domains are interlinked. Problem solving requires medical knowledge to make an appropriate judgment. Problem-solving skills can vary from day to day in an individual and means that a single observation may not give the true level of attainment. Different specialties will put different importance on different areas. Practical skills are more important for surgeons than occupational physicians. It is important that type and balance of assessments are appropriate (Crossley et al, 2002). We need to examine each area of the curriculum by a number of different assessments but based on a simple competency framework mapped to a curriculum.

Measurement

Measurement in educational assessments has a long history and methods are available to look at its characteristics (Ebel and Frisbie, 1991). At the minimum every assessment should be fair, valid, reliable (consistent and repeatable) and feasible. There are appropriate statistical

methods for both norm and criterion-based assessments (Ebel and Frisbie, 1991). Assessments must be made up of solid constructs and have both concurrent and predictive validity. There are problem around any assessment with error and reliability and these come from three sources: the assessment itself, the candidate and the assessor. A multiple choice question may be poorly constructed, the candidate may have personal factors impairing performance and the assessor may be untrained or may act inappropriately, for example. These can be minimized by careful construction of the assessment, and by standardization and training of the assessor. It is much more difficult to train and validate assessors in a work-based assessment compared with national assessments. Many trainers have difficulties in telling a trainee that their performance is below standard. This is a weakness of the RITA (record of in-service training assessment) process and is present in work-based assessments.

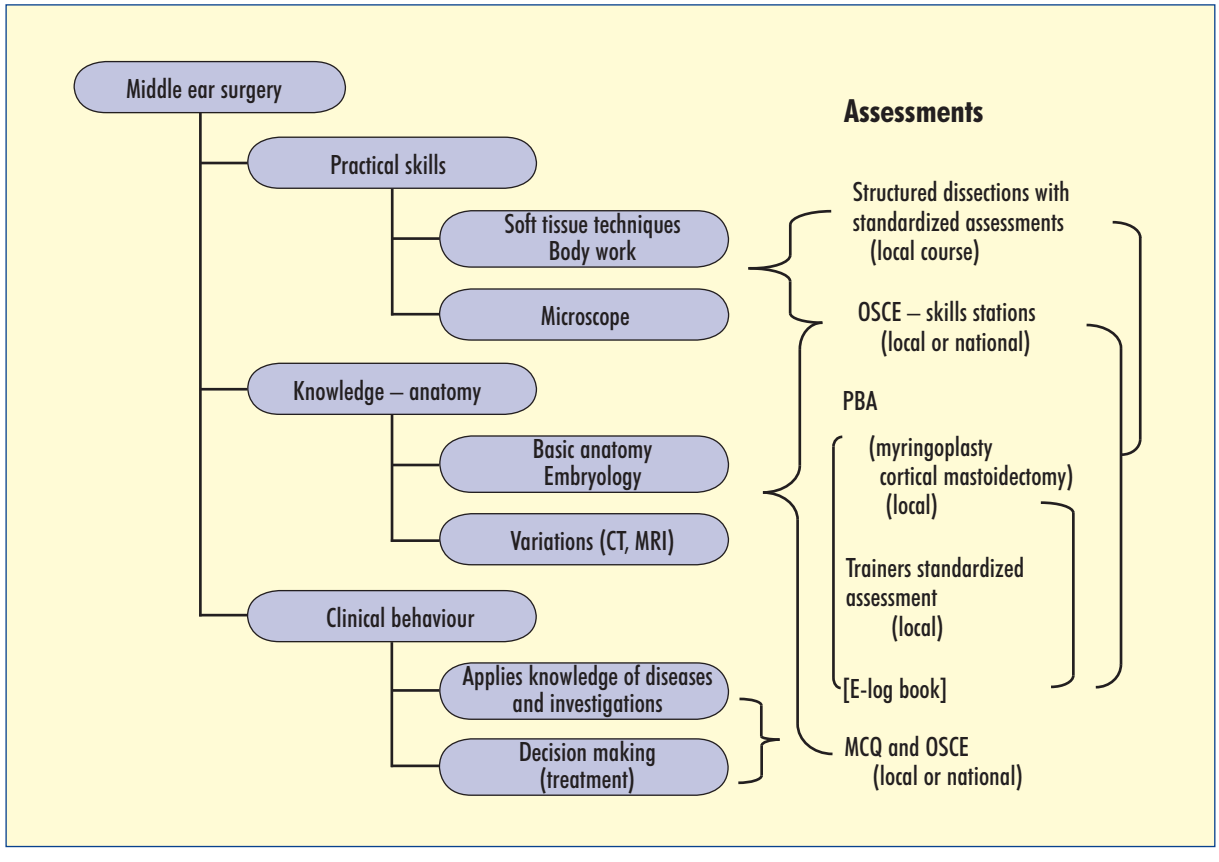
Administrative and construct problems with work-based assessments

Despite current practice, work-based assessments are not exempt from these standards. There is a long history of medical school assessment build on solid educational foundations, which could be used to validate these as the vast majority of doctors are appointed from UK medical schools.

The original article behind the mini-CEX (clinical exercise) assessment has the words 'a preliminary investigation' in the title (Norcini et al, 1995). This article demonstrates one of the problems with any work-based assessment – assessor compliance. Even though there were 397 usable assessments from 97 residents (registrar equivalent), nine of the assessments were rejected as these resident had completed only one assessment. The same rigours were not used on the assessors. There were also 97 assessors and 33 had completed one assessment and the median value was 2. This means that over half the assessors had marked two or less assessments in the year. Reproducibility required 10–12 encounters. The maximum number of assessments completed by one assessor was 26. This suggests that a few enthusiasts undertook most of the assessments. A more extensive study showed a similar proportion undertaking only one assessment (32%) (Norcini et al, 2003). These studies need to be validated against other assessments for reliability. This is particularly true for postgraduate training in the hospital specialties where there is a well recognized, if not reported, programme of national assessment.

Mini-CEX has another difficulty in its construction – the domain of professionalism, or humanistic skills as it is called in the USA. Crossley et al (2002) stated that professional behaviour is highly dependent on the nature

Figure 1. Mapping of a core area of practice through the three domains of practical skills, knowledge and clinical behaviour. CT = computed tomography; MCQ = multiple choice questions; MRI = magnetic resonance imaging; OSCE= objective structured clinical examination; PBA = procedure-based assessment.



and details of the problem in question. The domain of professionalism is firmly linked to that of a medical expert (Anonymous, 1996). This in turn is also related to that of scholar. Professionalism is a higher non-independent attribute of being both a medical expert and scholar. Does this category of professionalism pass the barrister test? When a litigious doctor who has failed to progress goes to court, his barrister asks you: 'What did you use to measure professionalism and was it valid?' To overcome this and other limitations of work-based assessments, longitudinal evaluation of performance in dental training has been advocated (Prescott et al, 2002). Even here professionalism is not well defined. We can measure medical knowledge, clinical behaviour, decision making and aspects of scholarship. If it was set up as part of a training programme, we could undertake a course on medical ethics and measure understanding afterwards. These are hard data and may be used to infer satisfactory competence.

The process

The processes for devising and setting assessments are clear. A curriculum must be developed and published, which is external to the programme, and it has to have well-defined objectives. This is much easier to do for postgraduate training after the foundation programme. If we start by using the Canadian Medical Education Directions for Specialist (CanMEDs) document, rank the importance of the seven roles or key attributes, and start with the most important one, medical expert, we may build up a curriculum and assessments around this (Anonymous, 1996). The attributes are:

1. Medical expert
2. Communicator
3. Collaborator
4. Manager
5. Health advocate
6. Scholar
7. Professional.

The doctor as a communicator falls into place besides medical expert but team working is difficult to pin down (Hays et al, 2002). Having defined our objectives we build a matrix or blue print of the curriculum to make sure that the curriculum is covered adequately (*Figure 1*). We devise our assessments in accordance with the framework. Some assessments are easier to deliver locally, other regionally or nationally. These assessments should have the same form wherever they are delivered in the UK. We must also consider the practicality of administering the assessments. Much of the current drive in assessment is politically driven and about power rather than the quality of the assessments: whoever controls the assessments controls training.

Medical graduates have had a great experience of assessments starting at school and held in a variety of places. All assessments are artificial whether they are taken locally or nationally, and held in the hospital, examination hall or tutorial room. There has been a trend to undertake work-based assessments recently on the assumption that they are more real and reduce anxiety. Any assessment will induce anxiety, but what is difficult to distinguish is whether the anxiety is produced by the place or the assessment. If there is no anxiety, those being assessed do not take the assessment seriously. There is concern about innovated school assessments as it is difficult to determine their operation and effect (Torrance, 1993). In school, it is recognized that teacher assessments and standardized assessment tasks have to be amalgamated and the feasibility of the development of higher order tests is debated. The last holds true in postgraduate education too.

Unlike school education, national examinations have been called into question. Most specialties have them in the first 3 years. Crossley et al (2002) stated that it was common practice to judge the competence of doctors for clinical practice exclusively on written evidence until recently, which only tested knowledge. This is incorrect. Most postgraduate exams in the UK have clinical components. There may be a need to change these assessments and make them more structured and fair but that is a different problem (Hutchinson et al, 2002). They could be changed into a competency format and rather than determine a threshold could be discriminating.

A model

Start with what works and develop from this. Each specialty will have different requirements dependent partly on their nature and partly on their size. The framework should be agreed nationally and standardized assessments used throughout the UK.

The Membership of the Royal College of Surgeons (MRCS) is common across all surgical specialties and is intercollegiate. It assesses a number of areas including knowledge, communication and clinical skills. Educationalists are employed by the Colleges to make sure that appropriate standards are met even if these are not reported (Hutchinson et al, 2002). This will change in the new ST1 and ST2 programme but the details have yet to be worked out. There is a move to drop clinical and communication skills as this is supposed to be testing the workplace.

There is an additional ear nose and throat (ENT) assessment, the Diploma in Otolaryngology and Head and Neck Surgery (DOHNS). It is set against a curriculum of 1 year of basic training and is open to all who wish to practice at this level. It consists of a written part made up from single best answers and extended matching questions, and a twenty-five station objective structured clinical examination (OSCE). There are seven clinical stations including communication, history taking and clinical examination, and there are eighteen stations that include case histories, investigations, equipment, gross and microscopic

pathology, medical knowledge, investigations and interpretation of data. It was set up with appropriate educational advice and was piloted. Both the MRCS and the DOHNS could be easily changed into this simple competency format. Any national exam such as the Membership of the Royal College of Physicians, which is fit for purpose, could do the same and so form the basis from which all other specialist registrar assessments are triangulated.

Subsequent assessments are built around a key objective, which is mapped to the three domains (*Figure 1*). Here it starts with a surgical skill, middle ear surgery. This approach could apply to many medical specialties too as they have practical procedures: stenting in cardiology, bronchoscopy in chest medicine and skin biopsies in dermatology. It could be organized around any areas that are considered key to a specialty. The figure is simplified for the sake of demonstration.

Most surgical specialties have electronic logbooks and these have summary sheets with indicator operations. Supervision levels range from 'observed' through to 'performed'. Summaries are no longer simple lists but have some interpretative value as they are linked to indicator operations, which are summarized year on year. These show increasing competence. There are many regional courses in surgery in orthopaedics and ENT, such as temporal bone courses. These could be assessed formally through the nationally agreed programme.

An OSCE could assess practical skills as well as medical knowledge, communication and other clinical areas. The ENT exit exam has a skills section in it. Some of these aspects may be assessed by multiple choice question examinations as well. These could be administered locally or nationally and as often as deemed necessary.

Work-based assessments would be included too but they must be workable. Procedure-based assessments have been developed in orthopaedics but may have some of the problems encountered in work-based assessments as we have found in our pilot in ENT. Simpler ones could be developed such as structured trainer reports. These could be given several times a year.

All these could be included in the portfolio together with other assessments.

Conclusions

National exams are well recognized by the public and could be changed into a simple competency framework. All subsequent assessments could be triangulated to this. A nationally agreed programme is required in each specialty with common assessments. OSCEs and multiple choice questions could contain appropriate stations in all specialties at the desired level. Work-based methods should be simple to use and could be triangulated to other assessments. These will require much more work to be fit for purpose. **BJHM**

Conflict of interest: none.

Anonymous (1996) *Skills for the new millennium: report of the societal needs working group.* The Royal College of Physicians and Surgeons of Canada, Ontario: 21

Crossley J, Humphris G, Jolly B (2002) Assessing health professionals *Med Educ* **36**(9): 800–4

Ebel RL, Frisbie DA (1991) *Essentials of Educational Measurement.* 5th edn. Prentice Hall, Englewood Cliffs: 370

Hays RB, Davies HA, Beard JD et al (2002) Selecting performance assessment methods for experienced physicians. *Med Educ* **36**(10): 910–17

Hutchinson L, Aitken P, Hayes T (2002) Are medical postgraduate certification processes valid? A systematic review of the published evidence. *Med Educ* **36**(1): 73–91

McClelland DC (1973) Testing for competence rather than for 'intelligence'. *Am Psychol* **28:** 1–14

Miller G (1990) The assessment of clinical skills / competence / performance. *Acad Med* **65**(suppl): S63–S67

Norcini JJ, Blank LL, Arnold GK, Kimball HR (1995) The Mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med* **123**(10): 795–9

Norcini JJ, Blank LL, Duffy FD, Fortna GS (2003) The Mini-CEX: a method for assessing clinical skills. *Ann Intern Med* **138**(6): 476–81

Prescott L, Norcini J, McKinlay P, Rennie J (2002) Racing the challenges of competency-based assessment of postgraduate dental training: Longitudinal Evaluation of Performance (LEP). *Med Educ* **36:** 92–7

Schuwirth L, Cees P, van der Vleuten M (2006) Challenges for educationalists. *BMJ* **333:** 544–6

Torrance H (1993) Combining measurement-driven instruction with authentic assessment: some initial observations if nation assessment in England and Wales. *Educational Evaluation and Policy Analysis* **15**(1): 81–90

KEY POINTS

- Competency is developed out of educational measurement.
- Competency assessment must fit into a defined curriculum with explicit objectives.
- Measurement in assessment is well developed.
- National assessments could be changed into a competency framework.
- Work-based assessments must be triangulated with established and validated assessments.