

Workplace-based assessment

Workplace-based assessment is now widespread throughout medicine. If carried out well, such assessments reconnect teaching and testing to the benefit of the learner. But workplace-based assessment brings a unique set of challenges to medical education and requires fresh thinking about how we consider and construct assessment programmes.

This article outlines some of the principles underpinning the design of workplace-based assessment and considers some of the tools that have been adopted for use within assessment programmes. The unique challenges of workplace-based assessment are considered, in particular the thorny issue of 'reliability'.

What is workplace-based assessment?

Workplace-based assessment refers to the assessment of what doctors actually do in practice and is predominantly carried out in the workplace itself. Workplace-based assessment in the training context relies on the use of tools for gathering information about aspects of trainees' work which are then used as vehicles for offering direct, timely and relevant feedback. The collection of workplace-based assessment data is learner-led and brought together, usually in a portfolio of evidence, to inform judgments about the trainee's overall progress.

So how does workplace-based assessment fit with traditional forms of testing in medicine?

Miller (1990) provides a useful pyramidal model (*Figure 1*) for mapping assessment methods currently available in medical education and illustrates how workplace-based assessment relates to the assessment of clinical competence.

'Knows' forms the base of Miller's pyramid, the entry point in the development of expertise. This tier is best assessed using simple knowledge tests such as multiple choice questions. The next tier up 'knows

how' seeks to measure understanding or application of knowledge and is assessed using instruments such as unfolding patient management problems, extended matching or short essay questions. Higher up, objective structured clinical examinations assess at the 'shows how' level where students are required to demonstrate not only knowledge and understanding, but that they can bring together and manipulate relevant knowledge, skills and attitudes in a controlled situation.

The problem is that what doctors do in controlled assessment situations correlates poorly with their actual performance in professional practice (Rethans et al, 2002). Assessment of competence in a contextual vacuum is all very well but how can we know what happens in the messiness of real professional practice – what the doctor actually 'does'? This is where workplace-based assessment comes into its own.

Is it useful?

The utility, or usefulness, of an assessment has been defined as a product of its reliability, validity, cost-effectiveness, acceptability and educational impact (van der Vleuten, 1996). Utility can be applied to an entire assessment system or to an individual assessment method or component of the system. The concept is important in that no single element should be regarded

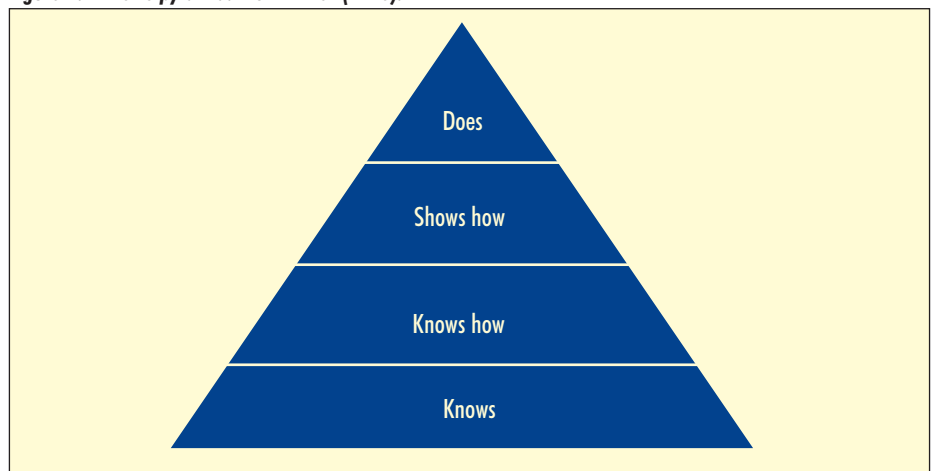
as predominant. Assessment design then inevitably leads to a trade off between individual elements. Thus, traditional approaches to maximize the reliability or reproducibility of assessments can have a negative educational impact on the learner by reducing the opportunity for meaningful developmental feedback. Workplace-based assessments offer high educational impact but might not be as reliable as other highly structured tests such as multiple choice questions.

Historically, the seductiveness of standardized testing led medical education to rely on externally administered assessments delivered at the end of programmes of training. Workplace-based assessment offers an opportunity to re-evaluate this situation and reintegrate teaching, learning and assessment (*Figure 2*), in other words, providing assessment that is 'built in' and not 'bolt on'.

From methods to programmes

Traditional approaches to medical assessment have been founded on the notion that domains of competence (e.g. problem solving, communication skills) are stable and generic. It was considered possible to design tests that assessed these domains separately and reliably leading to a 'one trait, one instrument' approach (Schuwirth and van der Vleuten, 2004). However,

Figure 1. Miller's pyramid. From Miller (1990).



Dr Tim Swanwick is Faculty Development Lead, London Deanery, London WC1B 5DN, Visiting Fellow, Institute of Education, London University, and Visiting Professor, University of Bedfordshire and **Dr Nav Chana** is Senior Lecturer in the Faculty of Medicine and Biomedical Sciences, St George's University of London, and Associate Director of General Practice, London Deanery

Correspondence to: Dr T Swanwick

there has been a growing realization that competence is specific to particular clinical situations or contexts. In order to overcome this problem, it is vital to sample widely across both the content of the curriculum and the contexts in clinical care is delivered.

Given the complexity of assessing professional competence it is now recognized that assessment should be construed as a programme of activity requiring the acquisition of quantitative and qualitative information from different sources. As a major contribution to such programmes, assessing doctors in their actual working environment offers the opportunity to gather information using a variety of different tools, so building a 'rich picture' of their working practices.

Workplace-based assessments will not replace standardized assessments. There are issues in relation to reliability as a result of inconsistent application of tools by different raters or assessors. There is potential conflict in the role of the trainer who is supervising the learner, but also involved in the assessment process. And there are problems of attribution when routinely collected clinical practice data are assessed. So in order to gain the benefits while mitigating the risks, a number of key issues should be considered in the design and implementation of such assessment programmes.

What to assess?

The areas chosen to assess in workplace-based assessment are usually expressed as a series of competencies. These should be blueprinted against the curriculum and, in the way they are expressed, should encourage learner development. Let us look at those three issues in a little more detail:

Competency-based

Workplace-based assessment is usually competency-based. Despite criticisms of competency-based education as a whole (Talbot, 2004), concerns have usually been voiced where competencies are viewed as narrow, reductionist and overly simplistic. Competencies used for designing workplace-based assessments are best written as holistic statements which are framed as 'a complex structuring of attributes needed for intelligent performance in specific situations' (Gonczi, 1994).

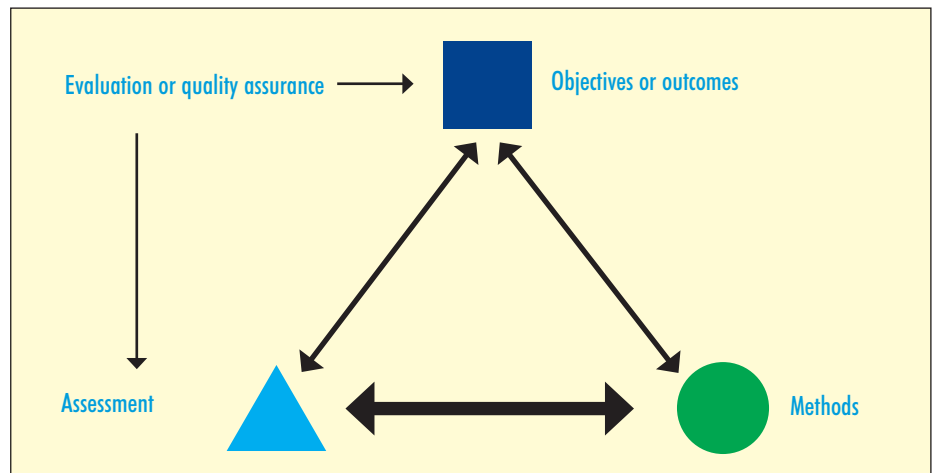


Figure 2. The educational paradigm: integrating teaching, learning and assessment.

Blueprinted

To ensure that assessments are integrated with the curriculum, competencies chosen for assessment should map directly onto the curriculum to ensure that there is both adequate coverage and widespread sampling. Some aspects of a curriculum will be more efficiently assessed through other means, clinical knowledge being an obvious case in point, however, some will be best assessed in the workplace. Indeed many aspects of professional performance such as team working, leadership and commitment to continuing professional development, are virtually impossible to assess in any other way.

Developmental

As already discussed, workplace-based assessment offers the opportunity to connect teaching, learning and assessment, and the developmental aspect of the assessment should therefore be a key feature. Developmental progressions in the literature, such as the novice to expert progression described by Dreyfus and Dreyfus (1986), may be helpful in constructing a developmental continuum of competence. Such a continuum has the advantage of explicitly illustrating the direction of travel for trainees, rather than merely pointing out the level below which they should not fall. This supports the concept of ongoing evidence collection throughout the training period, but with regular, well-circumscribed staging reviews at which the developmental framework is reviewed and the learner's progress through it judged.

So, workplace-based assessment provides useful formative and developmental

feedback but it also has a summative role and informs judgments about overall progress. This raises the tension of potentially mixing formative and summative elements, but it is possible to address this through the careful design of the assessment system. Separating the interpretation of evidence from its elicitation is one way around the problem (William and Black, 1996). In other words, when it is assessment time, the learner needs to know, and be adequately prepared for it.

How much evidence is enough?

Collecting 'sufficient' evidence is essential in making a judgment about the attainment of competence. As we have seen, sampling widely across a number of clinical and contextual situations is important to overcome the problem of case specificity. In the assessment of 'work' there is no single method that will do it all and a variety of sources of information will be needed. This gives rise to the notion of a 'tool-box' of assessment methods.

In considering individual tools it is worth recognizing that, even unstandardized, they can be made sufficiently reliable, provided the tools are used sensibly and expertly, and enough sampling occurs (van der Vleuten and Schuwirth, 2005). But it is important to remember that the tools themselves only form a small part of an overall assessment programme and attention should focus on the utility of the entire programme of assessment, not just the individual tools themselves.

Confidence in the reproducibility of judgments made on the basis of work-

place-based assessment can be improved through triangulation. This involves using a range of different methods to collect evidence using multiple raters over a sustained period of time. Triangulation with other assessments external to the workplace is also important and an overarching assessment strategy for each training programme, in which workplace-based assessment is supported by other test methods – such as those of ‘knowledge’ and ‘skills for clinical method’, is essential.

Which methods?

The methods for used for providing feedback and gathering workplace evidence in current use tend to be variations on one of four themes; observations of clinical activities, discussion of clinical cases, analysis of performance data and multi-source feedback.

Observations of clinical activities

Traditionally, clinical skills have been assessed by the ‘long case’ presentation. The problem of case specificity using this technique, limiting the potential to sample widely, has given rise to the mini-clinical evaluation exercise or mini-CEX (Norcini et al, 1995). This tool has been developed to assess the clinical skills that trainees most often use in real patient encounters. It is based on assessment of multiple complete or partial clinical encounters observed by an educational supervisor or other clinician.

The direct observation of procedural skills (DOPS) is another widely used tool, and one of a number of similar instruments based around the assessment of real-life activities where the focus is on the skill with which the activity was performed. ‘The consistent feature is that one or more assessors, who are trained in the assessment of that skill, make a judgment about a real life performance’ (Postgraduate Medical Education and Training Board, 2007).

A raft of other observational tools encompassing a wide range of workplace activities are in also current use including the procedure-based assessment of the Intercollegiate Surgical Curriculum, the mini-imaging interpretation exercise of the Royal College of Radiologists and the assessment of teaching of the Royal College of Psychiatrists.

Discussion of clinical cases

The origin of the use of case-based discussion in UK training assessment systems stemmed from their use in the General Medical Council’s performance procedures (Southgate et al, 2001) deriving originally from chart-stimulated recall oral assessments used in the USA and Canada. Case-based discussion is one of the evidence gathering tools used in workplace-based assessment in the UK foundation programme and is also being used in specialty training programmes such as in medicine, paediatrics and general practice.

Analysis of performance data

Norcini (2003) describes the basis for making a judgment on clinical performance data as having three potential sources; outcomes, process and volume. Outcomes of care, while being the most desirable measure, are limited by problems of attribution (to the individual), complexity, case mix and numbers. This is a particular problem in the assessment of trainee performance.

The process of care is more directly attributable to the individual doctor but effective processes do not necessarily mirror the best patient outcomes. The use of volumes of activity is premised on the basis that the more of a given activity that a doctor performs, the better their quality of care is likely to be. This basis for judgment is typified by the log books of the craft specialties such as surgery.

Multi-source feedback

The aim of using multi-source feedback to assess doctors in the workplace is to view a person’s work from a variety of perspectives. In medical settings, physician colleagues (peers), co-workers and patients can be asked to complete surveys about the doctor. The person being assessed receives feedback based on his/her own aggregate ratings, usually along with average ratings of others being assessed at the same time. There is also a clear opportunity for comparing self-assessment data with those provided by raters.

Multi-source feedback tools can be subdivided into peer-rating tools, such as the mini-PAT (mini peer-rating assessment tool) used in foundation training, and patient satisfaction questionnaires, a significant number of which are in use in the UK (Chisholm and Askham, 2006).

Portfolios

Workplace-based assessments are usually collected within a structured portfolio. A portfolio comprises a dossier of evidence collected over time, which demonstrates a doctor’s education and practice achievements (Wilkinson et al, 2002). There are many portfolio models (Webb et al, 2002) but in essence, if well constructed, a portfolio should chronicle the journey of a learner towards the attainment of professional expertise. A portfolio:

- Aims to serve as the reflective learning log of the learner, available to be shared with his/her educational supervisor
- Demonstrates the learner’s progress towards covering the breadth and depth of the curriculum
- Acts as a repository for assessments
- Provides a framework for learning agreements between learners and teachers
- Charts a learner’s progression and can help in making career choices and decisions.

The majority of portfolios used in medical education are web-based although with significant differences in structure and design between specialties and stage of training.

Quality assurance

Returning to the concept of utility, workplace-based assessment has huge strengths in the area of validity by virtue of its assessment of real or authentic material. Potentially it may have significant educational impact because of the reconnection of teaching and learning. Acceptability and cost-effectiveness are also potential winners but depend largely on how programmes are implemented. There are, however, significant issues with reliability as understood by traditional psychometric approaches. As Southgate et al (2001) point out, ‘establishing the reliability of assessments of performance in the workplace is difficult because they rely on expert judgements of unstandardised material’.

In workplace-based assessment there are several specific threats to reliability:

- Inter-observer variation: the tendency for one observer to mark consistently higher or lower than another
- Intra-observer variation: variation in an observer’s performance for no apparent reason (the ‘good day/bad day’ phenomenon)

- Case specificity: variation in the candidate's performance from one challenge to another, even when they seem to test the same attribute.

In the context of workplace-based assessment it is therefore helpful to reframe reliability as an attempt to maximize 'consistency and comparability'. Baker et al (1992) propose a number of activities that can help to do this, namely:

- Specification of standards, criteria, scoring guides
- Calibration of assessors and moderators
- Moderation of results, particularly those on the borderline
- Training of assessors, with retraining where necessary
- Verification and audit through the collection of assessment data.

It is clear, then, that the implementation of a successful workplace-based assessment programme will require training for assessors, arrangements for calibration, a procedure for the moderation of results and a raft of quality control checks. The more that teachers can be engaged in assessment, for example in selecting methodologies, generating standards and discussing criteria, the more the educational benefits of this powerful form of assessment can be realized.

texts using multiple raters; and processes in place by which evidence can be collated, synthesized and judged at regular intervals by an educational supervisor to assess the learner's progress with consistency and comparability across assessment programmes maximized through a robust programme of quality assurance. **BJHM**

Conflict of interest: none.

Baker E, O'Neil H, Linn R (1992) Policy and validity prospects for performance-based assessment. *Am Psychol* **48**(12): 1210–18

Chisholm A, Askham J (2006) *What Do You Think of Your Doctor? A review of questionnaires for gathering patients' feedback about their doctor.* Picker Institute, Europe

Dreyfus H, Dreyfus S (1986) *Mind over machine. The Power of Human Intuition Expertise in the Era of the Computer.* Basil Blackwell, Oxford

Goncz A (1994) Competency based assessment in the professions in Australia. *Assessment in Education* **1**(1): 27–44

Miller G (1990) The assessment of clinical skills/competence/performance. *Acad Med* **65**(Suppl): S63–7

Norcini J (2003) ABC of learning and teaching in medicine. Work based assessment. *BMJ* **326**: 753–5

Norcini J, Blank L, Arnold G, Kimball H (1995) The mini-CEX: a preliminary investigation. *Ann Intern Med* **125**: 795–9

Postgraduate Medical Education and Training Board (2007) *Developing and Maintaining an Assessment System – a guide to good practice.* Postgraduate Medical Education and Training Board, London

Rethans J, Norcini J, Baron-Maldonado M, Blackmore D, Jolly B, La Duca T (2002) The relationship between competence and performance: implications for assessing practice performance. *Med Educ* **36**: 901–9

Southgate L, Cox J, David T et al (2001) The assessment of poorly performing doctors: the development of the assessment programmes for the General Medical Council's Performance Procedures. *Med Educ* **35**(Suppl 1): 2–8

Schuwirth L, van der Vleuten C (2004) Changing education, changing assessment, changing research. *Med Educ* **38**: 805–12

Talbot M (2004) Monkey see, monkey do: a critique of the competency model in graduate medical education. *Med Educ* **38**: 1–7

van der Vleuten C (1996) The assessment of professional competence: developments, research and practical implications. *Advances in Health Science Education* **1**: 41–67

van der Vleuten C, Schuwirth L (2005) Assessing professional competence: from methods to programmes. *Med Educ* **39**: 309–17

Webb C, Gray M, Jasper M, Miller C, McMullan M, Scholes J (2002) Models of portfolios. *Med Educ* **36**(10): 897–8

Wilkinson TJ, Challis M, Hobma SO, Newble DI, Parboosingh JT, Sibbald JG, Wakeford R (2002) The use of portfolios for assessment of the competence and performance of doctors in practice. *Med Educ* **36**: 918–24

Wiliam D, Black P (1996) Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *Br Educ Res J* **22**: 537–48

KEY POINTS

- Workplace-based assessment is now widespread across all specialities and all stages of training.
- Workplace-based assessment offers the opportunity to connect teaching, learning and assessment.
- Workplace-based assessment has a dual function of offering focussed and timely feedback to trainees as well as providing data to support more long range judgments about trainee progress.
- Workplace-based assessment requires new ways of thinking about reliability based on maximizing consistency and comparability.

Conclusions

Workplace-based assessment offers the opportunity to connect teaching, learning and assessment, provides a means for assessment of problematic areas that require evaluation of real performance in practice and is a useful component of an overall assessment programme. In order for its benefits to be realized there needs to be clarity about what is being assessed through the identification of holistically described professional competencies; attention given to the developmental nature of the assessment; a variety of assessment tools used to gather evidence from multiple clinical con-