

# Assessment in medical education: roles for clinical teachers

**Richard Hays**, School of Medicine, Keele University, Staffordshire, UK

**T**he assessment of learners is one of those perennial topics in medical education that leads to debate about the extent to which assessment rewards or undermines 'real' learning. Perspectives differ according to roles. Some educators wonder if formal assessment is even neces-

sary, as motivated learners will achieve their learning goals without it. Some clinical teachers prefer not to have any role in assessment because they fear that this might harm their relationships with learners. Learners focus on the capacity of assessment to make or break careers, and fear

being wrongly judged as having failed to reach required standards. Regulators fear that assessment may allow learners with serious deficiencies to pass, which might affect patient safety. Educational institutions view assessment as being essential, but many invest more in managing, rather than in

**Learners focus on the capacity of assessment to make or break careers**

## We should use assessment to drive learning in the right direction

developing, assessment, as a protection against an increasing number of appeals against assessment decisions.

However, most medical educators agree that assessment is an integral part of curriculum development, delivery and evaluation, needing both 'backroom' science to develop the most appropriate assessment programmes and a group of expert clinicians to make any necessary judgements. This paper briefly describes the current approach to assessment in medical education, aiming to encourage more clinical teachers to participate in its development and delivery.

### THE GOAL OF ASSESSMENT

All assessment programmes should embrace two firm principles. First, assessment must reward learners who achieve the intended learning outcome(s) of a particular course. This requires that the learning objectives reflect the tasks of the medical student or graduate at his or her particular stage of education, and for the assessment objectives to be the same as the learning objectives (see Figure 1). The degree of congruence should be evaluated as part of course quality assurance. Failure to achieve this congruence will result in students merely learning what is assessed, rather than

what is taught – a form of 'hidden curriculum' – and lead to poor assessment validity.

Second, assessment should ensure that those who proceed to the next stage have met the required standards of their previous stage of education. This means that there must be extra vigilance regarding learners with borderline assessment results, because these results have an in-built margin of error. Which of the following errors is worse, in your opinion: a borderline candidate, whose ability may be *better* than the results on exam day indicate, but who is compelled to resit the exam before passing; or a borderline candidate, whose ability may be *worse* than the results on exam day, but is allowed through to the next stage without further assessment? These are errors that can be reduced through improved assessment methodology. The stakes are higher when the decisions are about permission to provide unsupervised clinical services.

### THE PURPOSE OF ASSESSMENT

Most clinical teachers will be familiar with the terms formative and summative assessment. *Formative assessment* (feedback) is used in a positive way to guide learners to improve their knowledge, skills and behaviours – learners do not fail if they achieve

low scores; while *summative assessment* is used to make pass/fail decisions. Recently, these terms have been used a little differently, as the focus is now on the development of a comprehensive assessment programme that both rewards learning *and* makes the right decisions about learners.<sup>1</sup>

All assessment can provide feedback to learners, with carefully chosen and prudently timed assessment guiding decision-making. Learners should always be clear about the purpose of individual assessment activities, as their behaviour may vary, depending on how results are used. The emphasis is moving away from major end-of-training barriers towards the use of within-training assessment that contributes to the final pass/fail decisions.<sup>2</sup> Using this method, decisions are made based on information gathered over time, rather than on a single occasion, which makes for a more robust system. Clinical teachers have a very important role to play in comprehensive assessment programmes, as students spend increasing amounts of their time in clinical settings, and should be assessed there.

### THE SCOPE OF ASSESSMENT

In simple terms, assessment programmes should assess *all* the required attributes of students and graduates at their particular stage of education, rather than only the aspects that can be measured easily. For example, consider a course in which the curriculum is revised to include an emphasis on professionalism in addition to integrated basic and clinical sciences, and yet the assessment uses the same question bank as the old curriculum. The likely result is that learners will stick to the old curriculum and ignore the addition of professionalism – another kind of hidden curriculum.

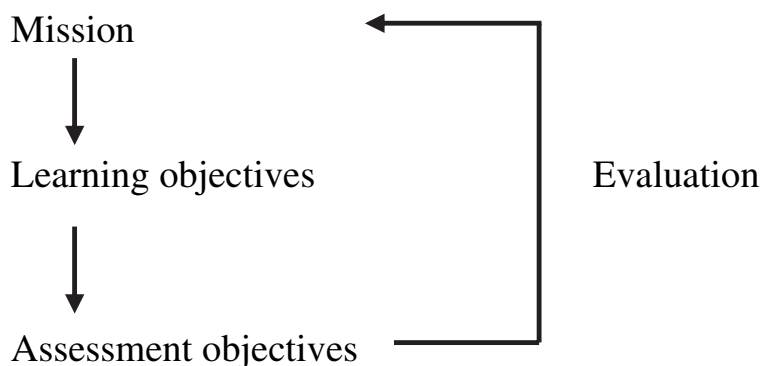


Figure 1. Achieving congruence in teaching and assessment.

If the saying ‘assessment drives learning’ is correct, and most agree that it is, then we should use assessment to drive learning in the right direction. All learning objectives should be included in the assessment, which will require them all to be expressed in more precise and measurable terms and included in the assessment blueprint. Not all of the learning objectives can be assessed in every assessment programme, so sampling is required, but more can be included if the entire assessment programme over the whole course samples items from the assessment blueprint.

Most current curricula are organised into ‘vertical’ conceptual groups labelled ‘domains’ or ‘themes’. These are designed to span the entire curriculum, and are labelled ‘basic sciences’, ‘applied clinical knowledge’, ‘clinical skills’, ‘ethical and professional behaviour’ and so on, and integrate the more traditional subjects or disciplines across the whole curriculum. Clinical teachers should engage actively in domain or theme groups, and make sure that the relevant material from their clinical discipline is integrated appropriately into teaching and assessment, not only across a single year, but throughout the whole course.

## THE LEVEL OF ASSESSMENT

Clinical teachers will generally be familiar with George Miller’s pyramid,<sup>3</sup> shown in Figure 2, which illustrates clearly the different levels of assessment. Use of the lowest level (simple knowledge) is diminishing, and should be confined to the earlier stages of the curriculum, when some basic building-block knowledge may be assessed. The majority of knowledge assessment at undergraduate level should be at the level of ‘Knows How’ and ‘Shows How’, the latter reflecting ‘Competence’. At postgraduate level, the majority of assessment should be at the level of ‘Does’. This is genuine, *in vivo* performance, which is what matters most in health care. Clinical teachers are often best placed to make these workplace assessments.

The recent trend in undergraduate assessment of including performance assessment wherever possible is both more valid and a better preparation for professional life. For example, examinations can show that students know enough and can do enough during exam week, but not that they can apply or maintain their knowledge, skills and behaviours subsequently. This is why some medical schools place their final examination barrier at the end of

the penultimate year of the course, and then use workplace assessment for the final year, during which students are constantly under the scrutiny of clinical teachers and can be judged on the application of what they have learned as they prepare for their first jobs. Clinical teachers should do most of the workplace assessment.

## ITEM WRITING

Assessment items are best written with the assistance of clinicians, who provide not only the clinical context to ensure that questions reflect credible health care issues, ideally based on real health care encounters, but also to provide correct answers. Writing high-quality questions is more difficult than it appears, and ‘correct’ answers may change as knowledge develops. Questions should reflect the tasks learners are expected to master, the wording should be precise and unambiguous, and the answers clear and agreed, ideally by a panel that includes representatives of different disciplines and specialties, if assessment is integrated. Ideally, item writing is supported by educators with the appropriate expertise to underpin the clinical relevance of clinical teachers’ questions with the necessary educational theory.

**Assessment items are best written with the assistance of clinicians**

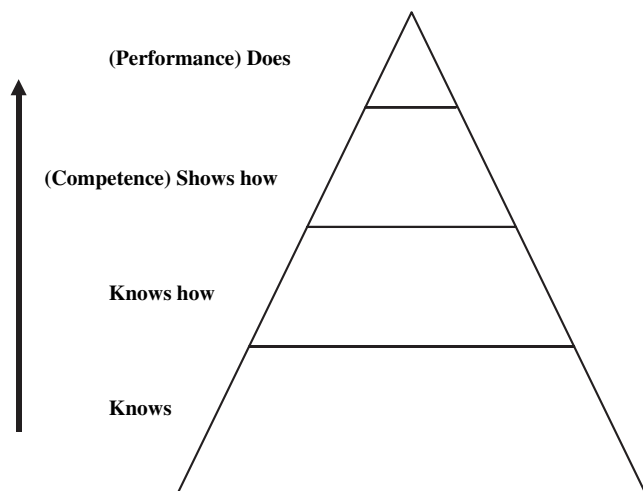


Figure 2. Miller’s pyramid<sup>3</sup>.

## STANDARD SETTING

Standard setting is the process by which the ‘pass’ score is determined. Many institutions have traditionally chosen an arbitrary pass score (often 50 per cent), an approach that is often little more than a guess. Some automatically fail a certain proportion of poor performers, or award honours to that proportion whose performance is excellent – an approach known as *norm-referencing*. A more modern approach is to try to determine pass scores more precisely, when a pass score

**In an integrated curriculum, a blueprint is particularly powerful**

may be more or less than 50 per cent, and may vary between different exams; this approach is known as *criterion-referencing*. This is a complex process that involves serious attention and the assistance of experienced educators, but also needs to include the views of clinical teachers, who understand clinical practice. And it is no longer enough for item writers to agree on correct answers; they must also have a sense of how borderline candidates might respond. More precise standards are then set, based on a combination of consensus and actual examinee performance data.<sup>4</sup> Commonly used methods include Ebel, Angoff and contrasting group methods, all described in more detail elsewhere.<sup>5</sup>

**ITEM BANKING AND BLUEPRINTING**

Assessment managers are always looking for more assessment items (questions), because they need to have available more quality-controlled items than will be used in any given year. Ideally, all items

are stored in some form of database (item bank), which is organised according to curriculum themes and content areas (a blueprint), and an assessment blueprint guides the construction of the whole assessment process.<sup>6</sup> In an integrated curriculum, a blueprint is particularly powerful, as it can ensure that all intended learning objectives of the course are assessed and, in doing so, facilitate evaluation of the course. Item banks need updating every year, so assessment writing is a never-ending process.

**ASSESSMENT METHODS**

Choosing particular assessment methods should not be done until the assessment tasks have been chosen from the learning/assessment blueprint – the reverse of what happens in many cases. A wide range of assessment methods is available, each with particular strengths and weaknesses, and methods should be chosen that best match the tasks being assessed. A simple example is that skill proficiency should be

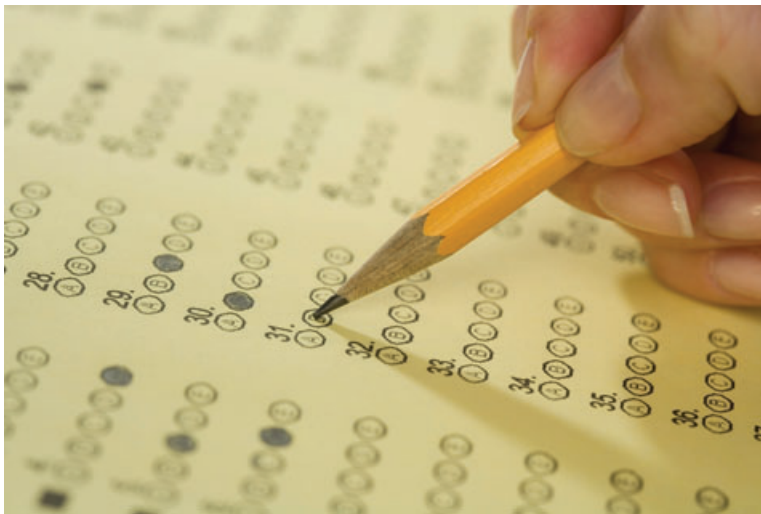
assessed by observation of the skill in action, not by a written or an oral description of how it might be performed. The important issue is not so much the individual methods themselves, but the utility of the assessment programme as a whole.<sup>1,7</sup>

Table 1 summarises common assessment methods, with a rough guide to their utility, although this also depends on the purpose and context of the assessment, as well as on the quality of assessment tools. New methods are emerging, particularly in performance assessment of senior medical students and junior hospital doctors, with better ways of assessing observed clinical encounters and procedures. A more comprehensive recent overview of individual assessment methods is available elsewhere.<sup>8</sup> Neither OSCEs nor learning portfolios are included in the table, because they are assessment formats that include several different assessment methods. Their design and scoring can be complex, and their challenge is deciding what to

**Table 1. The utility of common assessment methods**

Method	Validity	Reliability	Educational impact	Acceptability	Logistic ease
MCQ/EMQ	++	+++	++	+	+++
Essays	+	+	+	++	+
Short answer	+	+	++	+++	+
KFP	++	+++	+++	++	+
Case reports	+++	+	+++	+++	+
Vivas	++	+	+	++	+
Short cases	++	++	++	++	+
Long cases	++	+	+	++	+
Mini-CEX	++	++	++	++	+
DOPS	++	++	+++	++	+
Tutor ratings	+	++	+++	++	+
Video interactions	++	++	+++	+	+
Practice audit (referral, prescribing, patient records, MSF)	+++	+ to ++	+++	++	+

*Notes:* MCQ: multiple choice question; EMQ: extended match question; KFP: key feature problem; Mini-CEX: highly structured short case; DOPS: direct observation of practical skill; MSF: multi-source feedback.



include and how to combine their scores. All methods can be used for both undergraduate and postgraduate assessment, and both formative and summative assessment. The utility of assessment is best considered for a combination of methods, rather than individual ones. Reliability depends less on specific methods and more on the number of cases or items, and the total test time. In practice, most assessment methods are logistically complex, and all can contribute important information about learning progress, so methods should be chosen primarily on grounds of validity.<sup>9</sup> Clinical teachers can make valuable contributions to developing, delivering and marking assessments using all of these methods.

## PASS/FAIL DECISION AND RESITS

While many students and staff worry most about achieving or identifying distinctions and honours, the most important decision is in fact the borderline between pass and fail. Any pass mark is to some extent arbitrary (for example, the 50 per cent discussed earlier), although formal standard-setting procedures can improve the precision. The best way to define the borderline is arguably to use the Standard

Error of the Mean ( $S_E$ ), which is related to the reliability of the test. Hence a borderline score is the pass score  $\pm 1S_E$ , which defines scores about which there can be reasonable expectation of error.<sup>10</sup> The safest way to exclude or reduce the possibility of error is to ask students in this borderline zone to resit an equivalent standard examination. This is particularly important with higher-stakes examinations, such as those that confer eligibility for registration.

It is important to understand the real purpose of a resit. It should *not* be to allow learners with serious deficiencies to cram and 'scrape' a lucky pass, but rather to see if students with borderline scores achieve (or not) a similar score in two examinations. If performance improves to more than  $1S_E$  above the pass mark, then a pass decision is supported, while the same or a lower score supports a fail decision. Hence the role of assessors is to assess resit candidates against exactly the same standards as in the first examination. The answer to the question posed in the paragraph headed 'The goal of assessment' is that it should be better to retest and pass a student than to let one with problems go through without further testing and remediation.

## CONCLUSION

The development of assessment that supports learning in medical education requires both qualified, skilled educators who understand assessment methodology, and clinicians who understand what needs to be assessed and how to judge the performance of candidates. Clinicians play an essential role in item-writing and standard-setting procedures, as well as scoring student performance on examination days. More clinicians should feel comfortable about contributing to these tasks, and seek professional development in both developing assessment items and judging candidate performance.

## REFERENCES

1. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005;**39**:309–317.
2. Hays R, Wellard R. In-training assessment in postgraduate training for general practice. *Med Educ* 1998;**32**(5):507–513.
3. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;**65**(Suppl.9):S63–S67.
4. Norcini JJ. Setting standards on educational tests. *Med Educ* 2003; **37**:464–469.
5. Cusimano MD. Standard setting in medical education. *Acad Med* 1996;**71**(Suppl.10):S112–S120.
6. Hamdy H. Blueprinting for the assessment of health care professionals. *Clin Teach* 2006;**3**:175–179.
7. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education* 1996;**1**:41–67.
8. Epstein RM. Assessment in medical education. *New Engl J Med* 2007; **356**:387–396.
9. Schuwirth LW, Van der Vleuten CP. Challenges for educationalists. *BMJ* 2006;**33**:544–546.
10. Roberts C, Newble D, Jolly B, Reed M, Hampton K. Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Med Teach* 2006;**28**:535–543.

The most important decision is the borderline between pass and fail