# The postgraduate curriculum and assessment programme in psychiatry: the underlying principles

**Gareth Holsgrove, Amit Malik & Dinesh Bhugra**

**Gareth Holsgrove** is a professional medical educationalist, specialising in curriculum and assessment, as well as in training medical teachers and examiners. He was Medical Education Advisor to the Royal College of Psychiatrists and has been involved in developing the new College curriculum and assessment programme. Other previous roles include: Head of Postgraduate Educational Services at the Royal College of Psychiatrists; Head of the Academic Unit of Medical and Dental Education at St Bartholomew's and The Royal London hospitals; Director of Medical Education at the national medical school of the United Arab Emirates; Medical Education Consultant to several UK medical Royal Colleges and to the Aga Khan Foundation, the College of Physicians and Surgeons Pakistan and the World Health Organization; and Chair of the Postgraduate Medical Education and Training Board (PMETB) Assessment Approval Panel. **Amit Malik** is Consultant Psychiatrist at Hampshire Partnership NHS Trust. He is the lead for online implementation of workplace-based assessments (WPBAs) in psychiatry within the Royal College of Psychiatrists. He is the immediate past Chair of the Psychiatric Trainees' Committee at the Royal College of Psychiatrists and is President-Elect of the European Federation of Psychiatric Trainees. He is also a member of the Assessment Committee of PMETB. His research interests include service provision in liaison psychiatry and WPBAs in psychiatry. **Dinesh Bhugra** is Professor of Mental Health and Cultural Diversity at the Institute of Psychiatry, King's College London, and Honorary Consultant Psychiatrist at the South London and Maudsley Foundation NHS Trust. He is current President of the Royal College of Psychiatrists. His research interests include cultural psychiatry, psychosexual dysfunction and deviation.
**Correspondence** Dr Gareth Holsgrove, Gareth Holsgrove Ltd, 95 St Neots Road, Eaton Ford, St Neots, Cambridgeshire PE19 7AL, UK. Email: GarethHolsgrove@btconnect.com

**SUMMARY**

Assessment is key to the educational process and plays a significant role in looking at the progress trainees make as a result of training and personal development. Recent developments in curricula have led to substantial changes in assessing progress and attainment throughout postgraduate medical education in the UK. This article outlines the framework used to develop the postgraduate curriculum in psychiatry and describes the nature and purpose of the assessment programme that forms part of this new curriculum. The article considers the principles of medical education that are essential for the success of assessments, not only centrally in the development of the assessment system, but also locally in the delivery of these assessments. The overall context of developments in medical education, as well as the relationship between workplace-based assessments (WPBAs) and formal examinations, are described with specific references to developments in psychiatric training, its curriculum and assessments.

**DECLARATION OF INTEREST**

G.H. is Medical Education Advisor for the Royal College of Psychiatrists and chaired the Postgraduate Medical Education and Training Board Assessment Approval Panel. A.M. is the immediate past Chair of the Trainees' Committee of the Royal College of Psychiatrists. He has undertaken a literature review, funded by the College, into WPBAs. D.B. is President of the Royal College of Psychiatrists. The authors have all been involved in developing the assessment programme for postgraduate training in psychiatry.

Assessment is a key part of the educational process. It directs learning and significantly influences the learner's behaviour. Not only can assessment form the basis for planning educational programmes, it can also enable learners and their teachers to check the learners' progress and attainment. However, the process of assessment has potential pitfalls, which are mainly due to the content and methods of assessment, the expertise of the assessors, and the outcomes of assessment in respect to feedback and career progression. Issues connected with appeals procedures and feedback must form an integral part of the process so that both trainees and trainers/assessors can learn from the it. Another key problem is the burden of assessment and the extent to which this impairs, rather than supports, good learning practices and takes time away from actual learning.

In this article we summarise some characteristics of good practice in designing and carrying out assessments, and how the assessment programme relates to the curriculum and the learner's journey through it. However, it is helpful to set the context in which assessment in medicine is developing. For example, traditionally, 'assessment' has usually meant little more than formal examinations. However, assessments in the workplace (which might previously have been carried out occasionally and informally) are now becoming widely used in medical education. We also look at the relationship of examinations and workplace-based assessments (WPBAs) to the curriculum. We outline their contribution to the overall assessment programme and explain how information from both sources can be integrated to monitor progress.

## Recent developments in medical education

The past 20 years have seen significant developments in medical education in the UK and elsewhere. In the UK, these began mainly with changes in undergraduate education and came about with the introduction of new curricula following the recommendations in *Tomorrow's Doctors*, published by the General Medical Council (GMC) in 1993, and their implementation, which was facilitated by Kenneth Calman's Undergraduate Medical Curriculum Implementation Support Scheme (UMCISS). With the inception of the Postgraduate Medical Education and Training Board (PMETB) in 2003 (although it became fully functional only in 2005), many similar changes were introduced into postgraduate training. Supporting these changes, PMETB published documents that set the standards for curricula and assessments (Southgate 2004; Grant 2005). The most recent PMETB standards are summarised in Box 1.

It became clear that all assessments need to relate directly to the curriculum and that the assessment programme is integral to the curriculum. The PMETB expected postgraduate medical

## BOX 1  Summary of PMETB standards

**Curriculum purpose and development**

*Standard 1*  The purpose of the curriculum must be stated, including linkages to previous and subsequent stages of the trainees' training and education. The appropriateness of the stated curriculum to the stage of learning and to the specialty in question must be described.

**The assessment system must be fit for purpose**

*Standard 2*  The overall purpose of the assessment system must be documented and in the public domain.

**Content of the curriculum**

*Standard 3*  The curriculum must set out the general, professional, and specialty-specific content to be mastered, including:

- the acquisition of knowledge, skills, and attitudes demonstrated through behaviours, and expertise;
- the recommendations on the sequencing of learning and experience should be provided, if appropriate; and
- the general professional content should include a statement about how *Good Medical Practice* is to be addressed.

**The content of the assessment will be based on curricula for postgraduate training which themselves are referenced to *Good Medical Practice***

*Standard 4*  Assessments must systematically sample the entire content, appropriate to the stage of training, with reference to the common and important clinical problems that the trainee will encounter in the workplace and to the wider base of knowledge, skills and attitudes demonstrated through behaviours that doctors require.

**Managing curriculum implementation**

*Standard 5*  Indication should be given of how curriculum implementation will be managed and assured locally and within approved programmes.

**Model of learning**

*Standard 6*  The curriculum must describe the model of learning appropriate to the specialty and stage of training.

**Learning experiences**

*Standard 7*  Recommended learning experiences must be described which allow a diversity of methods covering, at a minimum:

- learning from practice;
- opportunities for concentrated practice in skills and procedures;
- learning with peers;
- learning in formal situations inside and outside the department;
- personal study; and
- specific trainer/supervisor inputs.

**Assessment system methods**

*Standard 8*  The choice of assessment method(s) should be appropriate to the content and purpose of that element of the curriculum.

**Supervision of the trainee**

*Standard 9*  Mechanisms for supervision of the trainee should be set out.

**Role of the assessor**

*Standard 10*  Assessors/examiners will be recruited against criteria for performing the tasks they undertake.

**Assessment feedback to the trainees**

*Standard 11*  Assessments must provide relevant feedback to the trainees.

**Standards for classification of trainees' performance/competence**

*Standard 12*  The methods used to set standards for classification of trainees' performance/competence must be transparent and in the public domain.

**Documentation will be standardised and accessible nationally**

*Standard 13*  Documentation will record the results and consequences of assessments and the trainee's progress through the assessment system.

**Curriculum review and updating**

*Standard 14*  Plans for curriculum review, including curriculum evaluation and monitoring, must be set out.

**Resources**

*Standard 15*  Resources and infrastructure will be available to support trainee learning and assessment at all levels (national, deanery and local education provider).

**Lay and patient involvement**

*Standard 16*  There will be lay and patient input in the development and implementation of assessments.

**Equality and diversity**

*Standard 17*  The curriculum should state its compliance with equal opportunities and anti-discriminatory practice.

(Postgraduate Medical Education and Training Board 2008)

education bodies to produce curricula for its approval in which the assessment programme was fully integrated. To meet this obligation, the Royal College of Psychiatrists developed an entirely new Core and General Curriculum, making it well placed to design the curriculum and its integrated assessment programme together. The College now has PMETB approval for both its curriculum and the supporting assessment programme.

## Mapping assessments to the curriculum

Mapping the assessment programme to the curriculum, which in turn is mapped to *Good Medical Practice* (General Medical Council 2006),

is a good way to satisfy PMETB Standard 4. Therefore, the Core and General module of the new curriculum was constructed using headings from *Good Medical Practice*. This model then served as a template upon which almost all of the specialty and subspecialty modules were subsequently structured.

### The curriculum framework

In the past decade there have been a number of national initiatives in the western hemisphere to define the roles (general categories of competencies) expected of doctors. Notable among these are the general competencies defined by the Accreditation

Council for Graduate Medical Education (ACGME) in the USA, the Royal College of Surgeons and Physicians in Canada's CanMEDS model and the GMC's *Good Medical Practice* guidance in the UK. These three models all set out the broad roles within which various specialties need to define competencies.

In the UK, PMETB required that all medical Royal Colleges could map their curriculum framework to the *Good Medical Practice* domains set out by the GMC. For the sake of simplicity, the Royal College of Psychiatrists' curriculum for specialist training is based directly on the *Good Medical Practice* framework. However, *Good Medical Practice* is quite complicated and not really intended as the basis for designing a curriculum, so mapping directly to this framework had some problems. In the original version of *Good Medical Practice* there are four main domains and the content is then organised in a hierarchical structure within each domain. Below the domain level is the subdomain, followed by major competency, aspect and finally supporting competencies.

Most of the major competencies have a number of aspects, and the supporting competencies for each aspect are set out under the headings of 'Knowledge', 'Skills' and 'Attitudes'. Table 1 shows just one branch from the first of the *Good Medical Practice* domains (Good clinical care) down to one of the supporting competencies, listed as 'Knowledge'.

Since the Core and General module runs throughout the 6 years of specialist training in psychiatry (ST1–ST6), it is necessary to indicate the developing competencies. This is because, in most instances, a specialty registrar (StR) in ST1 would be performing at a lower level of expertise than one approaching the end of their training. The College Curriculum Group decided to do this by placing the developing competencies in three categories ('Under supervision', 'Competent' and 'Mastery') and to use a colour code to indicate the stage of training in which different levels of performance should be achieved: red indicates specialist training stage 1 (ST1); gold, ST2 and ST3; violet, ST4 and ST5; and green, ST6. These colours are used not just in

the curriculum, but in all the documents relevant to each particular phase of training, such as the WPBA forms and descriptors.

It is important to note that competencies in the 'Under supervision' category become incorporated into the 'Competent' category at a later phase of training (usually the following phase), even if this has not been specifically stated.

## Developing an assessment programme

Many new assessment instruments have been developed as a result of the problems with the existing assessment processes. The number of students and trainees has grown exponentially, and this presented a problem in respect to conventional training and assessment methods because they were all predicated on an apprenticeship model (van der Vleuten 1996). There has also been an increasing emphasis on assessments taking place in the context of day-to-day practice. Furthermore, there is an increasing trend towards involving students directly and actively in their own education and assessments (Schuwirth 2004).

In developing an assessment programme that follows modern principles and is integrated into the curriculum, a variety of issues needs to be considered such as the purpose and psychometric properties of assessments, their blueprinting and their utility (Wass 2001). We consider each of these in the next three sections.

## Purpose of assessments

Each assessment should be considered in the context of its assessment programme, and PMETB requires that the purpose of an assessment be explicit (Southgate 2004). Assessments should be educational and formative (i.e. providing educational feedback) (Wass 2001), particularly since we know that assessment is the most important driver of learning (Newble 1983).

However, within a programme such as specialist medical training, assessments also need to have a summative or pass–fail function. The same assessment tool can often be used for both a summative and a formative purpose, but it is

**TABLE 1**   **Mapping curriculum competencies to aspects of *Good Medical Practice* domains**

| Curriculum hierarchy | Competency |
|---|---|
| Domain from *Good Medical Practice* | Providing good clinical care |
| Subdomain from *Good Medical Practice* | Providing a good standard of practice and care |
| Major competency | Undertaking clinical assessment of patients with mental health problems |
| Aspect of major competency | Consultation |
| Supporting competency (knowledge) | Psychiatrists apply knowledge of specific techniques and methods that facilitate effective and empathic communication between the psychiatrist, patient, carers, colleagues and the wider healthcare system, including: acknowledgement of diversity relating to age, gender, race, culture, disability, spirituality and sexuality |

very important that this should be made clear at the outset (Crossley 2002a). This is not easy in practice because competencies and categories of competencies overlap, and the most appropriate type and number of assessments need to form an assessment programme that assesses all the relevant competencies validly and, when they have a summative function, very reliably.

In the College curriculum, WPBAs have a predominantly formative function (discussed in greater detail later in this article). The MRCPsych examination, of course, forms the backbone of the summative assessments in postgraduate psychiatric training. The MRCPsych will remain mandatory for trainees to progress, complete their training and obtain their Certificate of Completion of Training (CCT).

## Blueprinting

As we have noted, assessment is the most powerful driver of learning in medical education. To a considerable extent, this is probably because trainees feel burdened by their workload and focus on learning only what is assessed. Therefore, it is reasonable to require that assessments validate the outcomes set by the curriculum. To achieve this, test content should be planned with reference to the learning objectives (the College uses the more recent Intended Learning Outcomes framework) – a process known as blueprinting (Wass 2001). A blueprint is a matrix in which the test designer determines how many items/tasks are to be assessed for each subject or category. Then all the outcomes to be measured are explicitly stated in the blueprint, thus allowing an assessment programme to be developed that contains and utilises appropriate types of assessment method in the varying clinical settings (Crossley 2002a). Inadequate blueprinting of assessments raises concerns about the validity of an assessment programme. The blueprint should ensure that appropriate forms of assessment are used to assess the various domains of the curricula (skills, knowledge, attitudes and so on; Wass 2001). The intricate relationship between the various aspects of clinical competence and the characteristics

of different assessment instruments means that multiple forms of tests should be used, particularly for high-stakes summative assessments.

Although not explicitly stated in the PMETB principles and standards documents (Southgate 2004; Grant 2005), there is a clear expectation that the assessment programme will feature WPBAs as well as formal, national examinations. In fact, now that all of the medical Royal Colleges have received PMETB approval for their assessments, it is clear that all have opted for such a combination, although the balance between workplace-based and formal assessments may vary considerably according to specialty and local circumstances.

Although the College has decided on a programme of WPBAs and formal examinations, over the next few years the blueprint of the assessments will need to be developed further.

However, the assessment programme has been designed in such a way that much of the curriculum can be assessed both in the workplace and in the MRCPsych. Table 2 shows the assessment matrix for 'Consultation', for which the place in the curriculum hierarchy was illustrated in Table 1. The assessment matrix shows that consultation skills can be assessed in the workplace using assessment of clinical expertise, case-based discussions, case presentations and the mini-Assessed Clinical Encounter. It can also be assessed in the MRCPsych in the objective structured clinical examination (OSCE).

## Utility of assessments

Utility has been defined as a multiplicative function of reliability, validity, educational impact, acceptability and cost, with different weights attributed to each (van der Vleuten 1996). As most of these elements cannot be quantified, this is a purely conceptual model and not a psychometric index. However, it does highlight the trade-offs involved in assessments, which are always necessary because perfect utility is a Utopian concept (van der Vleuten 1996). In reality, those responsible for the assessments must give different weights to the different component variables of utility, depending on the context and the purpose

**TABLE 2** Assessment of consultation skills

| Workplace-based assessment methods | | | | | | | | | MRCPsych assessment | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Assess-ment of clinical expertise (ACE) | Assessment of teaching (AoT) | Case-based discussion (CbD) | Case presentation (CP) | Direct observation of procedural skills (DOPS) | Journal club presentation (JCP) | Mini-Assessed Clinical Encounter (mini-ACE) | Mini-Peer Assessment Tool (mini-PAT) | Team assessment of behaviour (TAB) | Multiple choice questions (MCQs) | Extended matching questions (EMQs) | Objective structured clinical examination (OSCE) |
| ◆ | | ◆ | ◆ | | | ◆ | | | | | ◆ |

of the assessment (van der Vleuten 2005). In this model, the relationship between all of the variables has deliberately been kept multiplicative so that if one of the elements is zero then the utility will be zero. Let us consider each variable.

### Reliability

Reliability is the technical term that describes the extent to which the results of an assessment reflect all possible measurements of the same construct (Crossley 2002b). It is the property of assessment data that refers to how much the results of an assessment can be reproduced (van der Vleuten 2005). It is important because all stakeholders involved must have faith in its results. To achieve this, the results must be reproducible and are therefore likely to be reliable.

The internal consistency of an assessment is usually expressed as a coefficient (Cronbach's $\alpha$) with values ranging from 0 to 1. This is just one of the estimations of error, but it includes aspects of other error sources and can be calculated using SPSS software (the standard platform of the assessment analyst). This makes Cronbach's $\alpha$ convenient and it has proved very useful for many years to developers of tests; it still remains the most common contemporary measure. An $\alpha$-value of 0.8 is regarded as the minimum acceptable value, but this acceptability really depends on the purpose of the exam (van der Vleuten 2005). Generally speaking, the higher the stakes in an examination, the greater the reliability should be: $\alpha = 0.90$ is regarded as the gold standard for high-stakes examinations. In practical terms, however, because of their formative/summative characteristics and (if they are appropriately designed and utilised) high validity, a reliability coefficient below 0.8 would often be acceptable for WPBAs. This is an example of the utility trade-offs mentioned in the preceding section.

Moreover, work by Schuwirth and van der Vleuten, who are in the vanguard of assessment in medical education, is challenging our assumptions on assessments and the interpretation of the results. They question the value of relying solely on strict psychometric tools such as reliability and validity to interpret modern assessment methods such as WPBAs (van der Vleuten 2006). We feel that their work will lead to significant changes in assessment strategies within the next few years and we are prepared to develop the College assessment programme accordingly.

However, the current situation recognises that achieving good reliability in assessments in medical education poses two particular challenges. Crossley and colleagues discuss these in more

detail, pointing out that the professional role of a doctor comprises complex behaviour and is highly dependent on the characteristics of the problem at hand (Crossley 2002b: p. 92). The classical approach to statistical support for assessments, which includes the calculation of Cronbach's $\alpha$, calculates the components of reliability one by one and the most important of these components are summarised below. However, in recent years an extension of classical theory has become more prominent and Lee Cronbach himself has endorsed it (Cronbach 2004). Crossley and colleagues (2002b) introduce this generalisability theory particularly well. Essentially, the theory quantifies all the sources of error simultaneously. These sources include errors within and between the assessed and the assessors, as well as the random errors that occur in all assessments. Moreover, and of special interest to test developers, it utilises mathematical modelling to predict the generalisability coefficient ($G$) from a number of different simulations (such as changing the number and duration of assessments or using different numbers of assessors) based on pilot data.

The number of variables involved allows us to see that the reliability of WPBAs is subject to many challenges. These might include the following.

#### Interrater reliability

This refers to the extent that different assessors observing the same thing would make similar assessments. Some authorities claim that interrater reliability is the single most important component of reliability where direct ratings are used (as opposed, for example, to computer-marking in MCQs). Poor interrater reliability is a potentially serious problem in assessments – particularly in some oral examinations. However, interrater reliability can be improved to an often considerable extent by assessor training and by using structured assessment instruments. The crucial factor in assessing a doctor's competence is adequate sampling of their performance across different patients by different examiners. This has been found to have a greater impact on reliability than standardisation (van der Vleuten 2005). Therefore, the most straightforward way to increase interrater reliability is to use a reasonably large number of observers and patients/cases. This has obvious implications for feasibility and cost (Crossley 2002a), but reliability in examinations is never cheap.

#### Case specificity

This is also known as domain or content specificity and the reliability of assessment outcomes is more

dependent on what is being assessed rather than on how (van der Vleuten 1996), particularly if a number of cases are assessed by different assessors. This implies that a trainee may perform differently when assessed for the same competency in different clinical contexts or with different cases, thus affecting the reliability of clinical assessment. Just as with interrater reliability, the simplest way to overcome this problem is to increase the number of cases used to assess a competency; but this too is handicapped by feasibility and cost. Also, it is important to use predominantly cases of medium difficulty for the group of trainees being assessed (Downing 2004) because these tend to be the best discriminators.

### Intrarater reliability

This recognises that the same trainee, when assessed for the same competency in the same clinical context and by the same assessor, can perform differently on different occasions. This could be due to a variety of factors intrinsic to the trainee, the assessor or external factors. The reliability issues in this situation can, once again, be addressed by using multiple assessments that form the basis of the overall assessment of a competency. In other words, we can think of assessment as a mosaic, gradually building up a picture of progress and attainment, rather than the single snapshot over a short timescale that tends to occur in the examination hall.

These issues of reliability, along with systemic bias (e.g. gender, ethnicity, age) are important factors that affect reliability in assessment scores, yet their effects can be reduced by a number of strategies.

### *Validity*

This is a complex issue with many aspects. Put simply, it is a measure of how thoroughly, accurately and appropriately a test measures what it purports to measure (Brown 2006). Messick (1995) defines validity as 'the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other models of assessment'.

Reliability, of course, is a prerequisite for validity (if an assessment is not reliable it cannot be valid) and high reliability allows for a greater measure of validity. However, this does not mean that high reliability alone is sufficient to demonstrate validity (Streiner 2003). On the other hand, it is of course pointless to design a reliable assessment that has no validity.

Although a complicated matter, validity has traditionally been classified into five aspects: face,

content, construct, criterion and, more recently, consequential validity. Criterion validity is further divided into predictive and concurrent validity. However, the debate continues, so for our purposes we can say that, whatever the components and the balance between them, we can summarise validity to mean that we are reliably assessing the right things in the right way, using the right people and are having a positive effect on learning, behaviour, professional development and outcome.

This summary is probably adequate for practical purposes, but there have been criticisms of some of the technical aspects of validity. For example, Streiner & Norman (2003) discuss the concept of face validity ('Does a test appear to assess what it claims to?'). Another contemporary view sees all validity as construct validity, though this itself can be challenged. Nevertheless, whatever the components, validity can be summarised in technical terms as a process of hypothesis testing, wherein the aim of a validation study is to formulate a hypothesis about the inferences that can be drawn from the results of an assessment, and then collecting evidence to prove or disprove this hypothesis (Downing 2004).

In response to these challenges, the College aims to collect evidence of the validity of their assessment programme from a wide variety of sources and to use appropriate blueprinting of the curriculum's assessments to support the content validity of the assessment framework.

Predictive validity studies using longitudinal data (e.g. WPBAs predicting clinical and/or examination performance) will be invaluable when establishing the long-term credibility of the assessment framework. Focus groups, qualitative studies and survey questionnaires should also be considered in order to assess the consequential validity (educational impact) of the assessment framework.

### *Feasibility*

This is a particular issue with WPBAs and it needs to be evaluated in respect of the various clinical settings and of the number of assessors. More contemporarily, it must be evaluated with different types of assessor (e.g. senior doctors, other healthcare workers, patients, carers, simulated patients).

There are major and specific concerns regarding the feasibility of WPBAs, including such practical matters as assessor and trainee fatigue, yet it is also recognised that some trainees, in some specific areas of their work, require more assessment than others. To minimise potential disruption, a framework should be developed to give guidance

on how much assessment is enough to ensure confidence in the outcome of the training (the delivery and assessment process). Not only is there a danger of having too many assessments, it is also important that time is provided for assessments to take place. Above all, it has to be accepted that this is a huge cultural change in the delivery and assessment of postgraduate medical education and its overall assimilation into the in-service training system will take time to be properly implemented and researched. However, as the first generation of those being assessed in the workplace now becomes the assessors, the system should run more smoothly.

There are some other logistic issues to consider. Centralisation, especially in tasks such as assessor training, automating the processing, and reporting results, will lead to a more efficient system. Additionally, good local administration strategies are essential, including dissemination of information and the integration of assessments into day-to-day activities. Portfolios or similar tools should be used to collate assessments undertaken over time (Postgraduate Medical Education and Training Board 2006).

### Acceptability

For the assessment programme to be successful, the framework must be acceptable to all concerned – particularly the assessors and the trainees. Pilot data have shown that assessors might not properly administer a potentially highly valid and reliable assessment instrument if they are not convinced of its educational value. Neither are they likely to correctly use assessment instruments that significantly limit their freedom to employ their professional judgement.

The assessor's knowledge of educational research and the importance they believe it to have are known to be limited (van der Vleuten 1996). This is not particularly surprising, yet it needs to be addressed in the light of our findings about WPBAs. First, the assessors need to understand information about the relevance of a particular domain and the ability of the assessment tool to assess it effectively as part of the overall assessment framework.

Second, it is important for the test designers to listen to the feedback from the assessors and take this into account. This requires that the acceptability of the assessment instruments and the curriculum framework be evaluated on an ongoing basis; that the key concerns, principal benefits, reliability and validity be identified; and that this information be properly disseminated and acted upon.

### Educational impact

The educational impact of assessment is technically known as consequential validity and, although it has been recognised for several years, it is now becoming recognised as such an important factor that it merits further discussion in its own right as one of the fundamental issues in assessment.

Trainees who already feel burdened (maybe over-burdened) by the pressures of clinical work – being on-call, attending courses, participating in audits, giving presentations – are almost certain to concentrate their learning efforts on what is being assessed. Therefore, assessment can be a powerful driver of learning. In fact, 'assessment is usually the most powerful factor in the entire curriculum, because it determines the real curriculum, the one which the students follow, rather than the one which the faculty may intend or believe that they follow' (Holsgrove 1997a).

There are a variety of ways in which assessment can influence learning. These include the content (blueprinting), the format, the feedback, the scheduling (van der Vleuten 1996) and the consequences of failure (van der Vleuten 2006). Not only are students likely to learn topics that are assessed, they are more likely to learn and practise well those topics that are assessed more thoroughly or frequently, or to which the most importance is attached. For these same reasons, assessments can also have unexpected and unintended negative effects – for example, by focusing on trivia while ignoring the essentials.

To make the best use of assessment, it is best practice to use multiple formats of assessment within an assessment programme. It is important that each format is validated for a particular purpose. For example, MCQs have been traditionally used to test knowledge (typically, straightforward factual recall) but, when based on a clinical scenario, they can also be used to test clinical reasoning and the application of knowledge.

An important component of the educational impact of assessment is feedback. The PMETB requires that assessments provide relevant feedback to those being assessed (Southgate 2004) but it can also be extremely useful to provide feedback to the assessors. Therefore, feedback should be built into the assessment programme and should also link to action planning and the trainee's personal development plan. To achieve this, trainers should be trained to provide effective, formative and action-oriented feedback, which must include an assessment of the trainee's strengths and weaknesses, enable learner reaction, encourage self-assessment and help to develop an action plan.

### Cost-effectiveness

This is an important matter because good assessments are expensive to develop, deliver and quality assure. Assessor training takes time and money – and medical education in the UK is usually short of both. Therefore, it is important that adequate funds be identified at the beginning for development, implementation and quality assurance of the assessment programme. However, besides the development and quality assurance costs there are hidden costs such as the assessor's time, trainee's time and administrative costs. These should all be identified and made explicit.

There is also a broader question of who pays for assessments – trainees, trusts, deaneries, etc. (van der Vleuten 1996). In the present context, the College has borne the initial costs of developing many of the assessment instruments for both WPBAs and the MRCPsych examinations. This includes piloting them and developing the electronic portal for WPBAs.

## The Royal College of Psychiatrists' assessment programme

The College's assessment programme has been designed to determine or contribute to a number of different functions, all concerned with progression of a trainee towards achieving specialist registration as a psychiatrist.

### Purposes of the assessment programme

At the most basic, but extremely important, level, the assessment programme provides information to help the trainee and trainer to identify areas of strength and those aspects where further input and support are required. These latter aspects will be mainly identified through WPBAs, and this is why it is so important for trainees to take advantage of undertaking WPBAs early in each phase of training rather that succumbing to the temptation of leaving them all until the last minute – by which time it might be too late to rectify any shortcomings.

The programme of WPBAs will not only lead to eligibility to sit the MRCPsych examinations, but will help the trainees to prepare to succeed when reaching these important professional milestones.

Both the WPBAs and the MRCPsych will contribute to the annual review of competency progression (ARCP) process, which will determine whether trainees can proceed to the next stage of their training.

### Workplace-based assessments

Ten WPBA methods were identified and became the subject of literature reviews and practical experience in the pilot studies. These methods are discussed in detail in *Workplace-based Assessments in Psychiatry* (Bhugra 2007) and the pilot studies are also reported in the same book (Brittlebank 2007). Enlightened by findings from the pilot studies and further reflection and discussions, work on developing WPBAs continues and is likely to do so for perhaps the next 2 or 3 years. The function of WPBAs in the College curriculum is predominantly (but not exclusively) formative, to assist with planning educational programmes and to provide the feedback on progress and attainment that is essential to both trainees and their supervisors. However, there are also certain requirements for successfully completing WPBAs before progressing to the next stage of training or being eligible to take the MRCPsych. As mentioned above, the current eligibility criteria are available on the College website (www.rcpsych.ac.uk/exams/regulationsandcurricula.aspx).

### The MRCPsych

The MRCPsych, the backbone of summative assessment in the new curriculum, has been completely redesigned. Its content is determined by the curriculum, of course, and the methods have been selected according to three principles:

- they must supplement the WPBAs in sampling across the whole curriculum;
- they must have the high degree of reliability (accuracy and internal consistency) that contemporary best practice demands;
- they should be predominantly computer-marked to reduce the administrative workload (which for the 'old' MRCPsych was massive), to improve reliability by reducing the potential effects of different examiners awarding different marks for a similar standard of work, and to enable marks to be agreed and notified to candidates quickly after the examination.

The new examination comprises four elements (three written papers and one clinical practical) and uses just three examination methods. These methods are well established and thoroughly validated and have been discussed elsewhere (e.g. Holsgrove 1997b,c,d). The three written papers (Papers 1, 2 and 3) use a combination of MCQs using the single-best-answer format, and extended matching questions (EMQs; also called extended matching items – EMIs). Further details can be found on the College website (www.rcpsych.ac.uk/exams.aspx). On successfully passing all three written papers, candidates are permitted to sit the fourth part, the extended objective structured clinical examination (OSCE).

121

## Future developments

Although significant progress has been made by the College in gaining PMETB approval for its curriculum and assessment programme, there remains a great deal of work to be done over the next couple of years. The curriculum review mentioned above has now been completed and has been submitted for PMETB approval. However, considerable work still needs to be done regarding the development of WPBAs, particularly for higher specialist trainees. Existing WPBAs are being refined and new ones developed. In its further development and analysis of WPBAs, the College will be mindful of the risks of using these tools for entirely summative purposes, which might compromise the value of the assessment process and the feedback that follows it.

The MRCPsych will undoubtedly be fine-tuned on the basis of piloting, examiner experience and psychometric analysis. Therefore, it will be wise to see all this – the College curriculum and its programme of WPBAs and the MRCPsych examinations – as work in progress for at least the next 2 years.

## Acknowledgements

## References

Bhugra D, Malik A, Brown N (eds) (2007) *Workplace-Based Assessments in Psychiatry*. RCPsych Publications.

Brittlebank A (2007) Piloting workplace-based assessment in psychiatry. In *Workplace-Based Assessments in Psychiatry* (eds D Bhugra, A Malik, N Brown): pp 96–108. RCPsych Publications.

Brown N, Doshi M (2006) Assessing professional and clinical competence: the way forward. *Advances in Psychiatric Treatment*; **12**: 81–91.

Cronbach L (2004) My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*; **64**: 391–418.

Crossley J, Humphris G, Jolly B (2002a) Assessing health professionals. *Medical Education*; **36**: 800–4.

Crossley J, Davies H, Humphris G, et al (2002b) Generalisability: a key to unlock professional assessment. *Medical Education*; **36**: 972–8.

Downing SM (2003) Validity: on the meaningful interpretation of assessment data. *Medical Education*; **37**: 1–8.

Downing SM, Haladyna TM (2004) Validity threats: overcoming interference with proposed interpretations of assessment data. *Medical Education*; **38**: 327–33.

General Medical Council (1993) *Tomorrow's Doctors*. GMC.

General Medical Council (2006) *Good Medical Practice*. GMC.

Grant J, Fox S, Kumar N, et al (2005) *Standards for Curricula*. Postgraduate Medical Education and Training Board (http://www.pmetb.org.uk/fileadmin/user/QA/Curricula/Standards_for_Curricula_March_2005.pdf).

Holsgrove G (1997a) The purpose of assessing medical students. In *Teaching Medicine in the Community: A Guide for Undergraduate Education* (eds C Whitehouse, M Roland, P Campion): pp 179–82. Oxford University Press.

Holsgrove G (1997b) Principles of assessment. In *Teaching Medicine in the Community: A Guide for Undergraduate Education* (eds C Whitehouse, M Roland, P Campion): pp 183–5. Oxford University Press.

Holsgrove G (1997c) Assessing knowledge. In *Teaching Medicine in the Community: A Guide for Undergraduate Education* (eds C Whitehouse, M Roland, P Campion): pp 186–94. Oxford University Press.

Holsgrove G (1997d) Assessing clinical skills. In *Teaching Medicine in the Community: A Guide for Undergraduate Education* (eds C Whitehouse, M Roland, P Campion): pp 195–7. Oxford University Press.

Messick S (1995) Standards of validity and the validity of standards in performance assessment. *Educational Measure: Issues Practice*; **14**: 5–8.

Newble DI, Jaeger K (1983) The effect of assessments and examinations on the learning of medical students. *Medical Education*; **17**: 165–71.

Postgraduate Medical Education and Training Board (2006) *Quality Assurance, Quality Management and Assessment System Guidance (Revised)*. PMETB (http://www.pmetb.org.uk/fileadmin/user/QA/Assessment/QAQMASG_New.pdf).

Postgraduate Medical Education and Training Board (2008) *Standards for Curricula and Assessment Systems*. PMETB (http://www.pmetb.org.uk/fileadmin/user/Standards_Requirements/PMETB_Scas_July2008_Final.pdf).

Schuwirth LW, van der Vleuten CP (2004) Changing education, changing assessment, changing research? *Medical Education;* **3**: 805–12.

Southgate L, Grant J (2004) *Principles for an Assessment System for Postgraduate Medical Training*. Postgraduate Medical Education and Training Board (http://www.pmetb.org.uk/fileadmin/user/QA/Assessment/Principles_for_an_assessment_system_v3.pdf).

Streiner DL, Norman GR (2003) *Health Measurement Scales (3rd edn)*. Oxford University Press.

van der Vleuten CP (1996) The assessment of professional competence: Developments, research and practical implications. Advances. *Health Sciences Education*; **1**: 41–67.

van der Vleuten CP, Schuwirth LW (2005) Assessing professional competence: from methods to programmes. *Medical Education*; **39**: 309–17.

van der Vleuten CP, Schuwirth LW (2006) *How to Design a Useful Test: The Principles of Assessment. Understanding Medical Education*. Association for the Study of Medical Education.

Wass V, van der Vleuten CP, Shatzer J, et al (2001) Assessment of clinical competence. *Lancet*; **357**: 945–9.