CHAPTER 1

# Introduction: changes in training

Amit Malik, Dinesh Bhugra and Andrew Brittlebank

This chapter outlines the developments in postgraduate medical education in the UK that will influence psychiatric training for many years to come. It especially focuses on the role of the Postgraduate Medical Education and Training Board (PMETB), before its merger with the General Medical Council (GMC), and its development of principles for assessments in postgraduate training. There is a brief description of the Royal College of Psychiatrists' ('the College's') activities for the assessment of future trainees in the context of wider changes in postgraduate training in the UK. The challenges of the assessment of clinical competence and clinical performance are considered. Some of the basic concepts of competency- and performance-based assessments are outlined. Workplace-based assessments (WPBAs), tools for which are discussed in subsequent chapters, are placed in the context of familiar assessments and examinations of clinical competence and performance, including the traditional long case and Objective Structured Clinical Examinations (OSCEs). The concept of a programme of assessments is introduced, and there is mention of how these separate assessments may fit together for both formative and summative purposes. There is a section with some basic pointers that trainees and trainers must consider when undertaking WPBAs. Finally, there is a brief section on supervisor reports, which always have been, and will continue to be, indispensable in the assessment of trainee performance; these are discussed in greater detail in Chapter 9.

## Changes to training

Since the introduction of the PMETB and Modernising Medical Careers (MMC) in the early 2000s, many developments have influenced the structure of and principles underpinning postgraduate medical training in the UK.

### *Context*

The Postgraduate Medical Education and Training Board was set up as an independent regulator for postgraduate medical training and came live

in 2005. It established principles for curricula and assessment systems to guide postgraduate medical training in the UK. The Royal College of Psychiatrists' curricula were underpinned by these principles, thus making the system more workplace- and competency-based. Following the Tooke report (Independent Inquiry into MMC, 2008), PMETB merged with the GMC in April 2010.

## Modernising Medical Careers

In 2004, the four UK health departments published a policy statement outlining the planned reforms to the structure of postgraduate medical training, leading to the birth of Modernising Medical Careers. The two main components of these reforms were the foundation programme and the 'run-through' grade.

### *Foundation programme*

Since 2007, after graduating from medical school, doctors undertake an integrated 2-year foundation programme, which focuses on generic competencies and management of acute illness. The curriculum for the foundation years is competency-based and is assessed by a range of WPBAs. Most psychiatric training posts are in the second foundation year. There are some generic psychiatric competencies that all foundation trainees need to develop irrespective of whether they undertake a psychiatric placement.

### *'Run-through' grade*

The initial MMC proposals led to the development of a single run-through grade where specialty training had a single point of entry and there was to be no mid-point selection (from senior house officer to specialist registrar) as had traditionally occurred. Trainees, who are now called specialty registrars (StRs) were initially appointed annually to the new grade and continued, subject to satisfactory educational progress, till the completion of training and the attainment of the Certificate of Completion of Training (CCT) without any further mid-point selection. After this initial experience, psychiatric training has once again been decoupled and trainees after finishing their core training have to reapply to specialty-specific higher training programmes. All trainees in core and higher training in this StR grade have to train to a GMC-approved curriculum and are assessed according to a GMC-approved assessment framework.

## European Working Time Directive

The new European health and safety legislation, the European Working Time Directive (EWTD), has significantly reduced the amount of time doctors spend at work, as they now have to comply with a maximum of 48 hours' a week. The need to cover clinical services safely has led to widespread evolution of shift working, meaning that some of this 48-hour limit is not part of the traditional working day. This restriction on the amount of time doctors in training actually spend at work, being supervised by more senior

clinicians, is already affecting the training experience. This is primarily due to the fact that all medical training has traditionally been based on spending long hours at work and eventually experiencing various clinical situations. This will not happen with the reduced hours and a competency-based checklist has been proposed as a safety net, to ensure that trainees have achieved all essential competencies that will enable them, as consultants, to treat patients safely and competently. Additionally, a lot of the traditional teaching at the basic specialist (or core training) level is formal and classroom-based and this has already been affected by the new shift pattern of working. Finally, the reduced working hours, along with the development of functional teams (e.g. crisis resolution and home treatment, self-harm assessment), have further reduced the opportunities for trainees to see new patients first hand and to learn skills requisite for emergency psychiatry.

The College's commitment to involving patients and carers throughout specialist education in psychiatry reiterates the need for patient and carer involvement in further development and delivery of the curricula and assessment systems.

The run-through training grade went live in August 2007, supported by a national online selection system – MTAS (the Medical Training Application Service). The MTAS system was introduced in a cavalier manner leading to significant disruption to the lives, employment and training of junior doctors in the UK. The system was fatally flawed and its failure resulted in great discontent and fury among the medical profession, which ultimately led to the Secretary of State commissioning an independent inquiry, chaired by Sir John Tooke, to examine the framework and processes underpinning MMC. Among its many recommendations, the inquiry sought to move up the bar for professional ability from competence to excellence within postgraduate medical training. Following the Tooke report, psychiatric training was again decoupled and the run-through grade was broken down into the more familiar core and higher training structures. Additionally, the report also recommended the merger of PMETB and GMC, bringing all medical training, undergraduate and postgraduate, under the purview of one national regulator.

## *PMETB, GMC and Royal College of Psychiatrists*

From September 2005 to April 2010 PMETB was a sole statutory regulator of postgraduate medical education in the UK. It defined separate standards for assessments and curricula and set up a process of staged compliance with these standards from 2007 to 2010. To this end, the Royal College of Psychiatrists submitted new competency-based curricula to PMETB and following approval these were implemented for the trainee cohort commencing their training in August 2007. These curricula were based on *Good Medical Practice* domains of the GMC and were one of the first set of specialty curricula to be approved by PMETB. Following extensive feedback from trainees and trainers, these curricula were extensively rewritten

and mapped on the CanMEDS framework (Royal College of Physicians and Surgeons of Canada, 2005), which is extensively used in Canada for medical education and training and is felt to be more suitable for defining outcomes in medical education. In April 2010, PMETB merged with the GMC. Additionally, principles for assessment and curricula were subsumed into one document (General Medical Council, 2010) and the medical Royal Colleges were expected to submit curricula and assessment systems that complied with all GMC standards. The Royal College of Psychiatrists' revised curricula were approved by the GMC in June 2010 and have now been implemented for the training year commencing August 2010. There are essentially nine separate curricula, for six Certificate of Completion of Training (CCT) specialties and three subspecialties in general psychiatry, which have the core curriculum subsumed within each of them. Additionally, following the pilot feedback from 2006 (Chapter 13), in 2007 the College rolled out an assessment system which included WPBAs and examinations as an integral component of these curricula.

A major change for conventional assessment systems that existed pre-2007 is that the new assessment strategy relates to the entire training period, unlike in the past when assessments were undertaken only at discrete points in the form of periodically scheduled high-stake examinations. Clearly, the focus of postgraduate training is shifting away from simply gaining a certain number of marks or dichotomous pass/fail decisions in examinations, to national examinations being a vital component of a wider assessment system that includes WPBAs, educational supervisor reports and portfolio-based assessments. The details of the national examination undertaken by the College are discussed in Chapter 12.

## *Why are we interested in the assessment of clinical performance?*

Competency-based postgraduate training programmes in medical specialties are now part of many international postgraduate training systems, including those in the UK, the USA, Canada and now also Australia. The principles underlying the new training programmes ensure that there is emphasis on learning in practice, i.e. at the place of work, and that training and assessment revolves around the top two levels of Miller's pyramid for clinical assessment (Miller, 1990). Thus knowledge and its application will not suffice; it is not enough to 'know' or even to 'know how'. To 'show how' may reflect competency, but it is the apex of the pyramid that is of the greatest interest. Competency-based training begs the question of assessment of outcome at the 'does' level. This is the level of performance in daily clinical practice.

## *What is competence and what is performance?*

The fundamental components of clinical competence are knowledge, skills and attitudes. Competence in a clinical situation is the complex ability to

apply these three as needed according to the matter in hand. Performance is the enactment of competence. The assessment at the basic level relates to the questions 'Do they know it?' and 'Do they know how?'; at the competence level to 'Can they do it?'; and at the performance level to 'Do they show how?'

Unfortunately, things may not be that simple and most would agree that there is more to performance than an aggregation of competencies. What professionals do is far greater and more complex than the constituent parts that can be described in competency terms (Grant, 1999). Identifying a lack of competence may be easier than confirming attainment of a competency.

There have been valid concerns and criticisms of competency-based training as being reductionist (Talbot, 2004) and 'not fit for purpose' (Oyebode, 2009). However, McKimm (2010) quotes Gonczi defining competencies as 'a complex structuring of attributes needed for intelligent performance in specific situations', which more accurately reflects our aspirations for our future competent professionals. This definition, if used as the underpinning principle for competency-based training, should, in fact, enhance the standards of training and competencies towards excellence.

A cautionary note must be struck; four essential matters must be understood. The first is that there is no (current or future) single perfect tool for the assessment of overall clinical competence. Indeed, there are dangers in an endless pursuit of tools that break down competencies into even smaller assessable components, taking them further and further from the complexity of real clinical life. The second is that future direction is towards programmes of assessment in which different tools are employed. In this way performance, which includes the ability to apply a range of competencies in a professional setting, can be gauged. The third fundament is to consider the role of supervisor assessment. The supervisor is in a unique position to assess a trainee's day-to-day professional activities. Any programme of assessment of clinical performance must include this critically unique perspective and not just rely on numerical scores obtained from assessment tools. Finally, it is clear that ongoing evaluation and adjustment of the assessment programme will remain an essential component of its quality assurance process.

## *What should we be trying to achieve?*

With contemporary emphasis on competency-based curricula and assessment of performance at the place of service, great attention has been given to the development of a range of tools to meet the challenge of assessing clinical performance as described above in a valid, reliable and feasible fashion. Furthermore, there is a need to meet both formative and summative purposes of assessment, that is to provide feedback to trainees in an in-house training and developmental context and potentially to

provide data for the purpose of summary, such as informing eligibility for progress in training.

Although there are many methods for evaluating trainees' knowledge and some for measuring skills, the ability to reliably measure clinical performance is more limited. This ability is not contained in one instrument or method but in the concept and design of a programme of assessments adjusted in response to the changing nature of the relevant curricula. There is a choice of available instruments and methods that range from assessing what actually happens in the workplace and thus testing performance, through to the use of simulations, for example OSCEs that primarily assess competence, down to traditional written examination formats that assess knowledge and its application. These have been broadly categorised in Box 1.1.

**Box 1.1** Methods for the assessment of trainees' performance

1 Assessments of performance (what the doctor actually does in the service base):
- individual patient encounter, e.g. CEX (ACE), mini-CEX (mini-ACE)
- video of patient encounter in the workplace, e.g. as used in general practice for many years
- simultaneous actual patient encounter
- direct observation of a skill, e.g. DOPS or OSATS in obstetrics
- observation of team working
- multisource feedback, e.g. TAB, mini-PAT
- feedback from patients, e.g. patient satisfaction measures
- plus observation of performance in non-clinical areas, e.g. teaching, presentation.

2 Assessments of competence in simulated settings, including OSCE:
- consultation skills, e.g. with standard patient or other role-player
- discussion of clinical material, e.g. case-based discussion
- simulated practical procedure, e.g. on a mannequin or model
- simulated teamwork exercise
- critical thinking
- reflective practice, e.g. written-up case.

3 Cognitive assessments:
- knowledge, e.g. test such as MCQ, EMQ
- problem solving/application of knowledge, e.g. CRQ paper
- other written assessments.

ACE, Assessment of Clinical Expertise; CEX, Clinical Exaluation Exercise; CRQ, critical reading question; DOPS, Directly Observed Procedural Skills; EMQ, extended matching question; MCQ, multiple choice question; mini-PAT, Mini Peer-Assessment Tool; OSATS, Objective Structured Assessment of Technical Skills; OSCE, Objective Structured Clinical Examination; TAB, Team Assessment of Behaviour.

## Utility of assessments and assessment systems

In his seminal paper in 1996, van der Vleuten defined the concept of utility as a multiplicative product of reliability, validity, educational impact, cost and acceptability. *Reliability* refers to the reproducibility of results; for example, if the same trainee given the same examination repeatedly obtained the same score. *Validity* is a concept that describes if the test method is actually capable of measuring that which it purports to measure; for example, writing an essay on a mental state examination does not predict an individual's ability to perform such an examination. Another important consideration is *feasibility* (which relates to cost); although rigorous repetitive testing might give the answers closer to the truth in terms of competence, assessment and examination processes must be manageable within the constraints of time and resources available in the majority of clinical settings. Trainers and trainees have their own preconceived notions about various forms of testing and this and various other factors have an impact on the acceptability of an assessment programme. Without a significantly high *acceptability* by trainees and trainers, assessments cannot have long-term success. Finally, assessments drive learning and the content, format and programming of an assessment all contribute to its *educational impact* (van der Vleuten, 1996).

### *National exams or local assessments – what we know already*

A detailed overview of the literature on the various WPBA tools is presented in Chapter 2. This section briefly discusses some of the psychometric data around the traditional assessments of competence using examinations and puts this in the context of a few psychometric values obtained from studies on WPBAs. Standard settings in examinations are discussed in greater detail also in Chapter 2.

Examinations of clinical competence have traditionally used the long- and short-case *viva* approach. This approach has validity as candidates are assessed on real patients and asked problem-solving questions. However, as candidates are tested on different cases and judged by different examiners, reliability of the results may be flawed. Nevertheless, reliability for both can be markedly improved by increasing testing time (and thus sampling), for example from 1 to 8 hours. Reliability using the long-case examination has been estimated at 0.60 with 1 h of testing, increasing to 0.75 for 2 h, 0.86 for 4 h and finally 0.90 for 8 h of testing (Waas & Jolly, 2001). This finding has clear implications for the refinements of the Assessed Clinical Encounter (ACE) tool (see Chapter 5). To overcome the poor reliability of clinical examinations, objective clinical examinations were developed in the 1970s and have gained worldwide use. The Objective Structured Clinical Examination (OSCE) has become a familiar part of postgraduate examinations. However, its reliability is contingent upon careful sampling across clinical content and an appropriate number of stations, which generally means that several hours of testing time are in fact needed. For the OSCE,

reliability for testing times rises from 0.54 for 1 h to 0.82 for 4 h and 0.90 for 8 h of testing (van der Vleuten *et al*, 1988). For the mini-CEX, reliability commences at 0.73 for 1 h of testing and peaks at 0.96 for 8 h (Norcini *et al*, 2003).

**Standardised patients**

Another often used method for assessment of clinical competency is standardised patient examination. A standardised patient is a person who is trained to depict a patient in a way that is similar and reproducible for each encounter with different trainees. Hence they present in an identical way to each trainee. The standardised patient can be an actor, an asymptomatic patient or a real patient with stable, abnormal signs on examination. The advantages of using standardised patients are that the same clinical scenario is presented to each trainee (better reliability) and that clinical skills can be directly observed (higher face validity). Feedback can be instantaneous and can also be given from the point of view of the patient – the standardised patient would need to be trained to do this in a constructive manner. Using standardised patients has high face validity. Reliability varies from 0.41 to 0.85 (Holmboe & Hawkins, 1998). It increases with more cases, shorter testing times and less complex cases; it is better when assessing history-taking, examination and communication skills than when assessing clinical reasoning or problem-solving.

Standardised patients have been used in multistation exams such as OSCEs where trainees perform focused tasks at a series of stations. They have been used as a means of integrating the teaching and learning of interpersonal skills with technical skills and of giving direct feedback to trainees. By combining this with video-recording the student–patient encounter, there is a mechanism for the student to review the recording later as an aid to learning (Kneebone *et al*, 2005). This can be used as part of an assessment process and enables multiple raters to rate the trainee, thereby increasing the reliability.

A fundament of understanding with regard to these assessments is that sampling is the key factor in the determination of reliability rather than the degree of structuring and/or standardisation. This means that methods that are less structured and standardised, such as the Clinical Evaluation Exercise (CEX)/Assessment of Clinical Expertise (ACE) and mini-Clinical Evaluation Exercise (mini-CEX)/mini-Assessed Clinical Encounter (mini-ACE), can be almost as reliable as other more structured and objective methods. The finding also reinforces the need to develop and implement not merely single assessment tools but an overall schedule or programme of assessments.

## *Undertaking assessments locally*

The implementation of WPBAs in the postgraduate setting has been, and will continue to be, an incremental process. Regardless of the assessment

tool that is being used, the following pointers will help trainees and assessors in getting started.

**Pointers for assessors**

1 The assessor must assess the trainee for their particular stage of training.
2 The assessor will need to set out protected time to conduct the assessment.
3 It may be prudent to consider with the trainee in advance the sort of patient who will be at the appropriate level of complexity and challenge for a particular trainee's level.
4 It should be agreed in advance that an assessment will be formal, rather than the trainee or trainer mentioning this at the last minute before, or worse, during the assessment.
5 The competencies being assessed must be defined in advance and be appropriate to the situation that is to be observed.
6 The assessor should be fully familiar with the assessment form, the competencies being assessed and the associated performance descriptors.
7 The assessor should only assess the competency in question if they are capable of making a judgement about it, and they should only score observed competence.
8 Assessors must be trained in the use of the assessment tools and this should include performance-dimension training, training in observation-based assessment and a calibration exercise (Holmboe *et al*, 2004).

**Pointers for trainees**

1 The assessments should be trainee-led.
2 The trainee should have regular discussions with their educational supervisor about the competencies they need to attain during a period of their training and the type and number of assessments they could undertake to demonstrate the attainment of these competencies. These should be clearly included in a learning plan.
3 It might be prudent to undertake the initial assessment with the trainee's own educational supervisor, in order to fine-tune the learning plan for the next few months.
4 The trainee should also have discussions with their supervisor about the sort of case that would be appropriate for their stage of training.
5 The assessor should be given enough notice for the assessment to be set up, so that they can clear their schedule to facilitate an uninterrupted assessment.
6 The patient must give informed consent to participate in the assessment. This should be obtained by the trainee, recorded in the case notes and then reconfirmed in the presence of the trainer.

7 In early stages of training (core trainee (CT) year 1 and 2), it is entirely appropriate for the assessment to be undertaken by a higher trainee (specialty trainee (ST) year 5 or 6) or an experienced specialty doctor. In the latter stages of training, the assessments should be undertaken by a more experienced clinician, in order to provide feedback on higher-level competencies.
8 Analogously to an assessor, a trainee should also be fully familiar with the assessment form, the competencies being assessed and the associated performance descriptors.

## How do these assessments link together?

To answer this question, it is necessary to recall the purpose of assessment and then to consider these particular tools or forms of assessment. In overall terms, assessment is used for a number of purposes, such as: making judgements about the trainee's mastery of the knowledge, skills and attitudes set out in the curriculum; measuring improvement over time; identifying areas of difficulty; providing feedback; and planning future educational and professional needs. Attempts are often made to divide assessments artificially into formative and summative types, although in real life these functions of assessments overlap significantly. It is useful, however, to briefly revise what the two types of assessment mean.

### *Formative assessment*

A formative assessment is used to monitor a trainee's progress through a period of training. It involves using assessment information to feed back into the teaching and learning process. It should, and indeed must, foster learning and understanding.

The trainer (supervisor)–trainee relationship is fundamentally important to successful and effective formative assessments. Formative assessments must be built into the curriculum and not be added on as an afterthought. Observed clinical work is an excellent example of an assessment method used in formative assessment. However, its purpose is only realised when there is effective dialogue between the trainer and trainee. Hence the skills of supervising and giving effective feedback are as important for the prospective trainer/supervisor as any technical knowledge of the assessment tools themselves. For formative assessment to act as a means of improving competencies, both trainee and trainer must have a shared understanding of the assessment's position, power and purpose. Comparisons with a standard can be made and remedial action taken if required. The quality and honesty of the feedback is critical. A trainee cannot be told that they did well and then receive average ratings. Also, trainees who perform poorly should not be given high ratings. Such information will not assist in identifying strengths and weaknesses and thus will not enable the reshaping of educational objectives. It may also lead to an unsafe and

unsustainable clinical and educational relationship between the trainee and their supervisor as the trainee is allowed to work at stages beyond their real competence.

### *Summative assessment*

Summative assessments are usually undertaken at the end of a training course, module or programme and determine whether the educational objectives have been achieved. A grade or mark is given, indicating what has been learnt. Good summative assessment involves the analysis of evidence from as many sources as possible. In any form of summative assessment programme it is important that every essential aspect of the curriculum is covered to ensure that the resulting report validly reflects the trainee's ability.

In postgraduate training in psychiatry, summative assessments will provide a statement of achievement, at times serve as a guide to the continuation through the training grade (through the annual review of competence progression (ARCP) process) and will necessarily provide evidence to support the award of a certification of competence (such as the Certificate of Completion of Training from the GMC).

No single assessment is adequate to assess a trainee's overall competence. Experts have recommended a programmatic approach to developing assessments (Dijkstra *et al*, 2009). Various qualitative and quantitative methods to combine these assessments are described by educationalists, but these are beyond the scope of this book (for the interested readers, we recommend Schuwirth & van der Vleuten, 2006).

## Educational supervisor reports

Educational supervisor reports (further discussed in Chapter 9) are an overarching method used to assess overall performance in the working context and thus they have traditionally been associated with both high content and high context validity. They are discussed here because of their significance in a trainee's portfolio and their contribution to the summative assessment in the form of the ARCP. Ratings from a trainee's supervisor have been used for many years in local schedules of assessment. These have generally shifted from unstructured formats, such as letters which have low reliability and tend to be subjective, to more structured reports on specific areas of performance.

Supervisor reports are particularly useful for testing areas that are difficult to assess by conventional methods. These include personal attributes, attitudes, generic competencies and professional values (e.g. reliability), ability to work with others and time-keeping. Additionally, supervisors can draw upon evidence from more structured evidence around workplace-based assessments and well-structured feedback from colleagues and peers to support the report. Well-designed reports allow for assessment

against agreed standards and can identify underperforming trainees. Some utilise rating scales to assess various domains of a trainee's performance. Supervisor reports can be improved if supervisors are trained in their use, receive feedback on their reports and if multiple sources of evidence are used such as workplace-based assessment multisource feedback (see Chapter 6).

Supervisor reports must be designed with facility of use in mind and with an identification of the competencies to be assessed at a particular stage of training.

Finally, a debate has opened on the use of 'gut feeling' or trust in assessment (ten Cate, 2006). This moves beyond reliance on just structured and formal evidence of performance to an attempt to capture performance as a global outcome to expert judgement. This would be a more formal expression of a supervisor declaring who they would choose or trust to handle more complex clinical tasks or who they would be comfortable with treating a family member.

## Conclusions

Various government initiatives (including MMC) and changes in the legal frameworks (including the PMETB – now GMC – and EWTD) have transformed the delivery of postgraduate medical education in the UK. Notwithstanding this, the assessment of clinical performance has always been a complex task. The work of a doctor – the execution of their day-to-day clinical responsibilities – is more than just a sum of competencies. There is no single test that assesses this overall competence. Instead, what is required is a programme of assessments using different tools involving a range of clinical scenarios and settings and several assessors. The tools described in this book have potential to do just that provided they are employed as part of an overall assessment programme, with adequate sampling and triangulation through a range of assessors. These methods are at their most valuable when seen as educational tools that guide and mould learning, particularly the development of clinical skills. They can focus supervision, highlight progress, identify need and stimulate enquiry and understanding. Their development and implementation is fundamental to the delivery of the College's curriculum and thus to the development of psychiatrists of the future. Just as the curriculum itself will change in anticipation of and in response to both experience of its use in practice and new workforce needs, so these tools will be adapted and new tools will be developed.

The chapters that follow discuss the use of various assessment tools, the utility of portfolios in the future, the new national Royal College of Psychiatrists' exams and experiences from the WPBA pilot projects. Each chapter on assessment tools is based on the relevant background for each tool, discussions leading to their development, the description of the tools

along with the person descriptors, and the authors' early experience with the implementation of these tools. It is hoped that these details will help trainees, trainers and training programme organisers in this ever-changing world of postgraduate medical education.

## References

Dijkstra, J., van der Vleuten, C. P. M. & Schuwirth, L. W. T. (2009) A new framework for designing programmes of assessment. *Advances in Health Sciences Education*, doi: 10.1007/s10459-009-9205-z.

General Medical Council (2010) *Standards for Curricula and Assessment Systems – Revised*. GMC.

Grant, J. (1999) The incapacitating effects of competence: a critique. *Advances in Health Sciences Education*, **4**, 271–277.

Holmboe, E. S. & Hawkins, R.E. (1998) Methods for evaluating the clinical competence of residents in internal medicine: a review. *Annals of Internal Medicine*, **129**, 42–48.

Holmboe, E. S., Hawkins, R. E. & Huot, S. J. (2004) Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Annals of Internal Medicine*, **140**, 874–881.

Independent Inquiry into Modernising Medical Careers (2008) *Aspiring to Excellence: Findings and Final Recommendations of the Independent Inquiry into Modernising Medical Careers Led by Professor Sir John Tooke*. MMC Inquiry.

Kneebone, R. L., Kidd, J., Nestel, D., *et al* (2005) Blurring the boundaries: scenario-based simulation in a clinical setting. *Medical Education*, **39**, 580.

McKimm, J. (2010) Current trends in undergraduate medical education: teaching learning and assessment. *Samoa Medical Journal*, **2**, 38–44.

Miller, G. E. (1990) The assessment of clinical skills/competence/performance. *Academic Medicine*, **65**, 563–567.

Norcini, J. J., Blank, L. L., Duffy, D., *et al* (2003) The mini CEX: method for assessing clinical skills. *Annals of Internal Medicine*, **138**, 476–481.

Oyebode, F. (2009) Competence or excellence? Invited commentary on… Workplace-based assessments in Wessex and Wales. *The Psychiatrist*, **33**, 478–479.

Royal College of Physicians and Surgeons of Canada (2005) *The CanMEDS Physician Competency Framework*. Royal College of Physicians and Surgeons of Canada (http://rcpsc.medical.org/canmeds/index.php).

Schuwirth, L. W. T. & van der Vleuten, C. P. M. (2006) *How to Design a Useful Test: The Principle of Assessment*. Understanding Medical Education. Association of Medical Education.

Talbot, M. (2004) Monkey see, monkey do: a critique of the competency model in graduate medical education. *Medical Education*, **38**, 587–592.

ten Cate, O. (2006) Trust, competence, and the supervisor's role in postgraduate training. *BMJ*, **333**, 748–751.

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

van der Vleuten, C. P. M, van Luyk, S. J. & Swanson, D. B. (1988) Reliability (generalizability) of the Maastricht Skills Test. *Research in Medical Education*, **27**, 228–233.

Wass, V. & Jolly, B. (2001) Does observation add to the validity of the long case? *Medical Education*, **35**, 729–734.