

5. Objective structured clinical examinations (OSCEs):
- a have no proven reliability or validity in psychiatry examinations
  - b test only communication skills
  - c lend themselves easily to the testing of skills such as empathy or building rapport with the patient
  - d can be used to test ethics
  - e are difficult to set up.

#### MCQ answers

1	2	3	4	5
a F	a T	a T	a F	a F
b T	b T	b T	b F	b F
c F	c F	c T	c T	c F
d T	d T	d T	d F	d T
e F	e F	e T	e F	e F

## Commentary

Femi Oyebode

The introduction of objective structured clinical examinations (OSCEs) into the Royal College of Psychiatrists' membership (MRCPsych) examinations follows a comprehensive review and reform of College examinations. In responding to the paper by Wallace *et al* (2002, this issue), I will set out the context of the changes in the College examinations, identifying the justification for these changes, and respond to the specific queries raised by Wallace *et al*.

### Context

The aim of education is to ensure that students learn and know specific facts, comprehend the principles underpinning these facts, demonstrate the ability to analyse and evaluate the source of these facts and, furthermore, show an ability to synthesise information in order to produce new (that is, original) work. Assessments in their various forms attempt to test whether students can demonstrate mastery in these domains. For example, traditional multiple choice questions (MCQs) test for factual knowledge,

and newer MCQs in the form of extended matching items (EMIs) test for the application of factual knowledge to specific situations. In other words, the test methods are directed at specific domains.

In the College examinations, the critical review question paper tests the candidate's ability to analyse and evaluate information presented in research reports and the essay paper tests the candidate's ability to synthesise information and communicate it fluently in written format. In medicine, it is important also to test for competence in practical, clinical skills. This includes competence in particular performances such as interviewing the patient as well as competence in the application of knowledge to unique situations. The clinical examinations attempt to test mastery of skills and competence as well as application of knowledge.

### Methodological problems

It is true that all assessment methods have weaknesses. These weaknesses are all well rehearsed.

Femi Oyebode is Professor of Psychiatry at the University of Birmingham and Consultant Psychiatrist at the Queen Elizabeth Psychiatric Hospital (Mindelsohn Way, Edgbaston, Birmingham B15 2QZ, UK). He is also Chief Examiner of the Royal College of Psychiatrists. His research interests include the neuropsychology of psychotic symptoms and the application of ethical principles to the practice of psychiatry.

Multiple choice question papers have the advantage of being highly reliable but concerns remain as to their validity. However, the EMI format is recognised as being more likely to test clinical reasoning, that is, the application of clinical knowledge to practical situations. This is another way of saying that all tests continue to be refined with a view to reducing their acknowledged weaknesses.

Essay papers have the advantage that in-depth knowledge and understanding of a given area can be tested. But the scope for sampling a wide area of knowledge is markedly reduced as most essay papers require the candidates to write only one or two essays, whereas with MCQ papers a wide area of knowledge can be examined. The reliability of essay papers is low compared with MCQ papers. Thus, systems have to be found to improve the reliability of essays. Training of essay markers, regular auditing of marked scripts and double-marking as appropriate are methods widely used to reduce the potential variability in the marking of essays.

Before the introduction of the critical review paper the ability to analyse and evaluate data was untested in the College examinations. The methodology for this paper was invented specifically for this purpose. There are no significant issues regarding the content validity of the paper, but it could be argued that the terrain that it currently covers is narrow. The paper could be extended to include the appraisal of qualitative research data and data from other forms of scholarship, including the analysis and evaluation of books, essays and journalistic writings.

The traditional method of examining clinical skills is the single long case. This method has been adopted by medical schools and postgraduate professional examinations all the over the world since it was first introduced in Cambridge in 1842. As a method, it has face validity. A candidate sees a real patient, takes a history and conducts an examination of the patient with a view to discussing the case with examiners. The examination mirrors clinical reality and the candidate's competence is judged by senior colleagues. The problems with the traditional long case are self-evident. A candidate's competence is determined by his or her performance on a single case. In clinical practice we would be wary of reaching major decisions on the basis of a single case report and the argument goes that we should be equally wary of such decisions in clinical examinations. Furthermore, it is acknowledged that the outcome of the examination for the candidate can be adversely influenced by factors such as the difficulty of the case and the cooperation of the given patient. Examiner factors can also have undue influence: a single 'rogue' examiner can unfairly influence the outcome for the candidate. These concerns about the traditional long case are not

necessarily fatal to the long case as a method of examining candidates. However, we are obliged to find solutions to the problems described above. One solution to the traditional long case is the OSCE.

Objective structured clinical examinations allow a candidate's clinical competence over a relatively wide area to be sampled. Usually, in excess of 10 OSCE stations are used. In the College's OSCEs, there will be 12 stations. Undue influence of one examiner on the outcome for any candidate is much reduced. Thus, the reliability of the examination is markedly better than it is for the single long case. There are concerns, though, about face validity. Actors take the part of patients, complex clinical tasks are deconstructed into their component parts and it is the component parts that are tested. Thus, it could be argued that the skill to examine a single case and make sense of it in all its complexity is not examined. Furthermore, especially in general medicine and surgery, physical examination findings are often normal because simulated patients are normal people. This raises issues about the test's capacity to assess the candidate's ability not merely to conduct a sound physical examination but also to recognise and to describe abnormality.

What is clear from the discussion above is that a fair, reliable and valid examination will have to adopt a multiplicity of methods to assess candidates. The more methods, and convergence of results from these varying methods, the more confident we will be that the outcome for candidates is an accurate reflection of their knowledge and skills. This is the so-called method of triangulation, and it is the overall approach that the College takes. Thus, in the MRCPsych examinations, Part I candidates have to satisfy the examiners in a modified MCQ paper, an EMI paper and an OSCE. Part II candidates will have to satisfy the examiners in a modified MCQ paper, an EMI paper, a critical review paper, an essay paper, a traditional long case and a structured oral examination.

---

## Response

---

Wallace *et al* (2002) pose a number of questions directly about OSCEs. I will now consider each of these in turn.

### **In view of this being a 'high-stakes' postgraduate examination, how are reliability and validity being established?**

The validity of examinations can be determined in a number of ways. Usually, an examination must relate to a published curriculum. The new College

curriculum (Royal College of Psychiatrists, 2001) was developed following wide consultation within the College. This presupposes that any examination that follows this curriculum derives its legitimacy from it, that is, it validly examines what the body of the College has determined as requisite to the proper practice of psychiatry in the UK. There are ancillary methods for ensuring validity. A 'blueprint' for the preparation of the examination must be developed. A blueprint is simply a grid or map developed from the curriculum that allows questions to be commissioned or the examination itself to be audited. Such a blueprint exists for the College's OSCE. It is also possible to test for validity by carrying out a survey of clinicians and trainees to assess how far the questions reflect clinical realities in psychiatry. The College has every intention of carrying out such a survey when appropriate. Predictive validity, as opposed to content validity, can be tested by exploring how far performance on the OSCE predicts performance on other aspects of the examination or on career progression. Aspects of the predictive validity of the OSCE will be investigated in the future.

Objective structured clinical examinations are recognised as highly reliable examinations. Data collected from the College's pilot OSCE, yet to be reported, confirm this. The  $\kappa$  score for the examination as a whole was about 0.8. The second pilot OSCE took place in April 2002, and the statistic of the OSCEs will shortly be reported.

#### **What is the added value of OSCEs over the current method of examination?**

Objective structured clinical examinations have all the advantages discussed above. They allow a wide area of skills to be tested and reduce the impact of any one examiner on the overall outcome for the candidate. They also allow the College to test areas of practice currently unexamined. These include: the ability to communicate diagnosis and treatments to patients and their relatives; physical examinations; interpretation of results; and communicating complex judgements to other clinicians, including nurses, physicians and senior psychiatrists.

#### **What skills can be adequately tested by binary checklists?**

The marking schedule is not binary. The schedule is a 5-point scale as used in other College clinical examinations. OSCEs are used in the General Medical Council Professional Language and Assessment Board (PLAB) examinations and have

been introduced by the Royal College of Physicians, among others. There is no reason to think that there is any problem with marking OSCEs. OSCEs are objectively marked. This means that the weighting of particular objectives within each OSCE station is determined before the examination and the examiner's task is to award marks for each objective as listed on the mark sheet. Whether the candidate passes or not is determined by his or her performance on these objectives and by the relative weighting of these objectives.

#### **How can key psychiatric skills such as empathy and building rapport with patients be assessed?**

All the OSCE stations in the College examinations will have communication skills as an objective. In practice it is not difficult to identify the candidate who is unable to establish rapport and empathise with the patient. Inability to use all the well-known interviewing techniques such as open-to-closed cone questions, summary and reflective statements will indicate to examiners that this is a poor candidate. In the College's pilot OSCE held in April 2002, the reliability of the simulated patients' assessment of the candidates' communication skills was tested. In North America, simulated patients' opinions are taken into account in determining the candidate's final mark. There is no intention to do this in the College examinations. However, we will have data, following the next pilot examination, to explore this issue in more detail.

---

## **Conclusion**

---

The changes to the College examination, of which OSCEs are only an example, demonstrate the College's commitment to continue to improve its examinations in line with best current evidence. In many areas, the Royal College of Psychiatrists' approach is in advance of other medical Royal Colleges. The aim is to have the fairest, most valid and reliable examination that is possible.

---

## **References**

---

- Royal College of Psychiatrists (2001) *Curriculum for Basic Specialist Training and the MRCPsych Examination*. London: Royal College of Psychiatrists.
- Wallace, J., Rao, R. & Haslam, R. (2002) Simulated patients and objective structured clinical examinations: a review of their use. *Advances in Psychiatric Treatment*, **8**, 342-348.