Multiprofessional Faculty Development - Structured Assessments of Clinical
Long and short cases

In the traditional long case the holistic appraisal of the examinee’s ability to interact with, assess and manage a real patient is a laudable goal and contributes to the face validity of this method. However, in recent years there has been much criticism of this approach related to variations in examiner stringency, unstructured questioning and global marking without anchor statements and patient variability - in information disclosure, demeanour, comfort and health. Perhaps more importantly, examinees’ clinical skills may vary significantly across cases, so that assessing examinees on one patient cannot not provide generalizable estimates of a candidate’s overall ability (Norcini, 2001, 2002).

Although apparently similar to an OSCE in providing a larger range of short cases, the differences between short cases and an OSCE are that different students rarely see the same patients, cases differ in their complexity, the same two assessors examine the student at each case, the examination is not structured and the examiners are free to ask any questions they wish. These factors result in poor reliability of this method and OSCEs have superseded this genre of assessment.

See the module Workplace-based assessments for more information on different types of clinical assessments.
Objective structured clinical examinations

The OSCE is an assessment format in which the candidates rotate sequentially around a series of structured cases located in ‘stations’, at each of which specific tasks have to be performed, usually involving a clinical skill, such as history taking, examination of a patient or a practical skill. The marking scheme for each station is structured and determined in advance. There is a different examiner and a time limit for each station. The basic structure of an OSCE may be varied in timing for each station, use of checklist or rating scale for scoring, use of clinician or standardised patient as examiner, use of real patients or manikins, but the fundamental principle is that every candidate has to complete the same assignments in the same amount of time and is marked according to a structured marking schedule.

The use of OSCEs in the quantitative assessment of competence has become widespread in the field of undergraduate and postgraduate medical education since they were originally described (Harden and Gleeson, 1979). The main reasons are the high reliability of this assessment format and the equity that results from all candidates being presented with the same test. Some characteristics of a good OSCE are listed below.

What makes a good OSCE?

- **Blueprinting**: ensure the test content maps across the learning objectives of the course
- **Station development and piloting**: writing stations that function well
- **Examiner training**: engage the examiners, consistency of marking contributes to the reliability
- **Simulated patient training**: consistent performance ensures each candidate is presented with the same challenge
- **Organisation**: make detailed plans well in advance

Thinking point:
Taking each of the elements of ‘what makes a good OSCE’, which do you think will have the most impact on (a) the learners’ experience and (b) the teachers’ experience?

Reliability

Essentially the OSCE was developed to address the inherent unreliability of classical long and short cases. OSCEs are more reliable than unstructured observations in three main ways:

1. Structured marking schedules allow for more consistent scoring by examiners according to pre-determined criteria.
2. Candidates have to perform a number of different tasks across clinical, practical and communication skill domains – this wider sampling across different cases and skills results in a more reliable picture of a candidate’s overall competence.
3. As the candidates move through all the stations, each is examined by a number of different examiners, so multiple independent observations are collated. Individual examiner bias is thus attenuated.
To enhance reliability, it is better to have more stations with one examiner per station than fewer stations with two examiners per station (van der Vleuten and Swanson, 1990).

**Validity**

‘Content’ validity is determined by how well the sampling of skills matches the learning objectives of the course or degree for which that OSCE is designed (Downing, 2003). The best way to ensure an adequate spread of sampling is to use a blueprint.

**Educational impact**

The impact on students’ learning resulting from a testing process is sometimes referred to as ‘consequential’ validity. It is well recognized that students focus on their assessments rather than the learning objectives of the course. Explicit, clear learning objectives aligned with clinical skills assessment content and format can be a very effective way of encouraging students to learn the desired clinical competencies.

There is a danger in the use of detailed checklists as this may encourage students to memorize the steps in a checklist rather than learn and practice the skill. Rating scale marking schedules encourage students to learn and practice skills more holistically.
Blueprinting

Blueprinting is a powerful tool which helps to focus the OSCE designers on the exact nature of what they wish to test and to relate this to the learning outcomes (Table 2).

Table 2

How to construct an objective structured clinical examinations Blueprint

<table>
<thead>
<tr>
<th>Review the curriculum learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Decide on domains of skills to be tested</td>
</tr>
<tr>
<td>• Map the domains against the learning objectives</td>
</tr>
<tr>
<td>• Sampling: decide on the proportion of stations in each section</td>
</tr>
<tr>
<td>• Calculate your total testing time; ensure appropriate time is allowed for the task at each station</td>
</tr>
</tbody>
</table>

Once the blueprint or framework for an OSCE is agreed (Figure 1), the individual stations can domains be planned and classified according to this blueprint. This ensures adequate sampling across subject area and skill, in terms of numbers of stations covering each skill and the spread over the subjects/systems of the course being tested.

Figure 1  Example of an objective structured clinical examination blueprint

<table>
<thead>
<tr>
<th>System</th>
<th>History</th>
<th>Explanation</th>
<th>Examination</th>
<th>Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular</td>
<td>Chest pain</td>
<td>Discharge drugs</td>
<td>Cardiac</td>
<td>Blood pressure management</td>
</tr>
<tr>
<td>Respiratory</td>
<td>Haemoptysis</td>
<td>Respiratory</td>
<td>Peak flow</td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>Abdominal pain</td>
<td>Gastroscopy</td>
<td>Abdominal</td>
<td>Rectal examination</td>
</tr>
<tr>
<td>Reproductive</td>
<td>Amenorrhoea</td>
<td>Abnormal smear</td>
<td>Cervical smear</td>
<td></td>
</tr>
<tr>
<td>Nervous</td>
<td>Headache</td>
<td>Eyes</td>
<td>Ophthalmology</td>
<td></td>
</tr>
<tr>
<td>Musculo-skeletal</td>
<td>Backache</td>
<td>Hip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generic</td>
<td>Pre-op assessment</td>
<td>Consent for post-mortem</td>
<td>IV cannulation</td>
<td></td>
</tr>
</tbody>
</table>
Station development

It is important to write the station material well in advance of the examination date so that the stations can be reviewed and tried out prior to the actual assessment. Station material should include:

- Construct: a statement of what that station is supposedly testing, e.g. this station tests the candidate’s ability to examine the peripheral vascular system
- Clear instructions for the candidate: to inform the candidate exactly what task he/she should perform in that station
- Clear instructions for the examiners: to help the examiner at that station to understand his/her role and conduct the station properly. Include a copy of the candidate instructions
- List of equipment required
- Personnel requirements: Whether the station requires a real patient or a simulated patient and the details of such individuals (age, gender, ethnicity)
- Simulated patient scenario: if the station requires a particular role to be played
- Marking schedule: this should include the important aspects of the skill being tested, a scoring scheme for each item and how long the station should last. The marking schedule can be either in a checklist format or a rating scale (Figure 2). Items can be grouped into three broad categories: process skills (e.g. rapport, questioning and listening) content skills (e.g. appropriate medical or technical steps or aspects of the task or skill being tested) or management skills (appropriate set questions in specific relation to the case).

Figure 15.2 Global ratings vs checklist scores

When OSCEs were first introduced extensive detailed checklists of each step of a clinical task were produced for each station. Checklists often focused on easily measured aspects of the clinical encounter and the more subtle but critical factors in clinical performance were overlooked or ignored.

The use of rating scales to assess the performance of clinical skills has been shown to be reliable when used by expert examiners (Cohen et al. 1991). Examiner training can improve their reliability further (Hodges 2003).

It is more effective to use checklists to assess technical skills in the earlier stages of learning (at the ‘novice’ end of the learning trajectory) and to use rating scales to assess more complex skills, especially with increasing levels of professional competence (Arnold 2002; Hodges et al. 2002).

Thinking point:
How are patients used in the clinical assessments in which you have been involved?
Think about:
(a) Real patients
(b) Standardised patients
(c) Simulated patients
(d) Simulations

How might each of the various types of patient be used most appropriately in clinical assessments?
What are some of the advantages and disadvantages of involving each of the types of patient?
Simulated patients

It is best to use well-trained simulated patients for consistent performances in communication skills stations. It is desirable to have people across a range of ages and ethnicities as well as a balanced gender mix. Training and monitoring simulated patients is essential to ensure consistent performance - a significant factor in the reliability of the examination.
Real patients

Patients do not always give the same history repeatedly; they can become tired or unwell and may develop new signs and symptoms or even lose old clinical findings; however they can be a most valuable resource and need to be treated as such. Using real patients in OSCEs adds greatly to the validity of the assessment. Ideally patients should be used to assess the detection of common chronic clinical signs. For each clinical sign assessed you will need several patients and even the most stoical patient should not be expected to be examined by more than 10 students in the course of a day. Ideally patients should be swapped in and out of a station to allow them to have sufficient rest time.

See the Module Involving Patients in Clinical Teaching for further ideas on using different types of patients.
Examiners

OSCEs require large numbers of examiners: this can be strength, as candidates are observed and scored by clinicians, but also one of its potential weaknesses, as inconsistency between examiners will reduce the fairness and reliability of an OSCE.

Considerable resources should be devoted to examiner training. Training sessions should cover the principles of OSCEs and the role of examiners i.e. to assess and not to teach or conduct vivas, to adhere to marking schedules, and respect the role of the simulated patient. Training can usefully involve both the marking of pre-recorded or role-played OSCE stations, after which the marking is reviewed with the clinicians.
Practical considerations

The smooth running of OSCEs is highly dependent on the detail of the practical arrangements made in advance and it is worth putting some effort into this to ensure a satisfactory day of examinations. There are many aspects to consider and these are covered extensively in the Understanding Medical Education booklet by Boursicot et al (2007).
To sum up

Assessment of clinical competence is a crucial part of the basis on which decisions are made about the ability of clinicians and doctors in training. Any method of assessing clinical skills should be considered in the context of a wider programme of assessment, which should include the assessment of knowledge, clinical examination skills, practical procedure skills, doctor-patient communication, problem-solving ability, management skills, relationships with colleagues as well as professional attitudes and behaviour.

One of the most important aspects of assessing clinical skills is the range of sampling across a candidate’s skill base; this has to be taken into account when designing any assessment. OSCEs can assess clinical, communication and practical skills but are still situated in the context of an examination setting. To assess doctors in the context of their professional practice requires the use of different formats in the work place.

- Objective structured clinical examinations (OSCEs) are widespread in the assessment of clinical competence
- OSCEs are a fair and reliable method of assessing clinical skills
- OSCEs should be blueprinted to learning outcomes
- Developing high quality OSCE stations takes time and effort
- Training of simulated patients and examiners is essential
- Authenticity is important for test validity
Further Information

This module was written by Dr Katharine Boursicot, Reader in Medical Education, Head of Assessment, St Georges, University of London.

The author would like to acknowledge Trudie Roberts and Bill Burdick on the UME Understanding Medical Education booklet, OSCEs and other assessments of clinical competence, on which a large part of this module is based.

References

Boursicot KAM, Roberts TE, & Burdick WP (2007). OSCEs and other assessments of clinical competence ASME, Edinburgh
Self-Assessment Activities

Select one or more of the activities below to develop your skills in assessing educational and learning needs and setting educational objectives.

If you are registered on the site, you can write up your reflections in the reflections area. Click on the my area link at the top of the page to access your personal pages. Please note you must be logged in to do this.

1 Long and short cases

Consider one or two clinical assessments used in your educational context.

Would you consider that these used examples of long or short cases? If so, which of these?

Find examples of criteria and marking schedules used.

Using the criteria set out in the module, how effective do you think these are in ensuring reliability, validity and facilitating learning?

2 Blueprinting the OSCE

Taking an OSCE in which you are involved or of which you are aware, look at the test content and use the blueprint example in Figure 3 to analyse how this maps across the learning objectives or outcomes of the programme.

In which areas does the blueprint identify clear mapping to learning outcomes?

Are there any gaps or deficiencies in the blueprint?

What do you think are the reasons for this and what other assessments are being used to assess the relevant learning outcomes?

3 Planning and developing an OSCE

You have been asked to develop an OSCE for students/trainees learning with you in your speciality.

What issues would you need to address or questions you would need to have answered in each of the following areas:

(a) Developing stations
(b) Identifying and training examiners
(c) Designing scenarios
(d) Identifying and recruiting patients (real or simulated)
(e) Designing a marking schedule
(f) Standard setting
(g) Practical logistics

Based on your answers to the questions above and your likely role, what professional development needs have you identified and how will you take these forward?