



The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration

Kamran Z. Khan, Kathryn Gaunt, Sankaranarayanan Ramachandran & Piyush Pushkar

To cite this article: Kamran Z. Khan, Kathryn Gaunt, Sankaranarayanan Ramachandran & Piyush Pushkar (2013) The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration, Medical Teacher, 35:9, e1447-e1463, DOI: [10.3109/0142159X.2013.818635](https://doi.org/10.3109/0142159X.2013.818635)

To link to this article: <https://doi.org/10.3109/0142159X.2013.818635>



Published online: 22 Aug 2013.



Submit your article to this journal [↗](#)



Article views: 33563



View related articles [↗](#)



Citing articles: 76 View citing articles [↗](#)

WEB PAPER
AMEE GUIDE

The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration

KAMRAN Z. KHAN¹, KATHRYN GAUNT², SANKARANARAYANAN RAMACHANDRAN² & PIYUSH PUSHKAR³

¹Manchester Medical School, UK, ²Lancashire Teaching Hospitals NHS Trust, UK, ³Aintree University Hospitals Trust, UK

Abstract

The organisation, administration and running of a successful OSCE programme need considerable knowledge, experience and planning. Different teams looking after various aspects of OSCE need to work collaboratively for an effective question bank development, examiner training and standardised patients' training. Quality assurance is an ongoing process taking place throughout the OSCE cycle. In order for the OSCE to generate reliable results it is essential to pay attention to each and every element of quality assurance, as poorly standardised patients, untrained examiners, poor quality questions and inappropriate scoring rubrics each will affect the reliability of the OSCE. The validity will also be influenced if the questions are not realistic and mapped against the learning outcomes of the teaching programme. This part of the Guide addresses all these important issues in order to help the reader setup and quality assure their new or existing OSCE programmes.

Introduction

This Guide is the second in the series of two Guides on the OSCE. The first Guide focuses on the historical background and educational principles of the OSCE; knowledge and understanding of these educational principles is essential before embarking upon designing and administering an OSCE. We would advise the reader to familiarise themselves with the contents of Part I prior to reading this part of the Guide.

In this second part we aim to assist the reader in applying the theoretical knowledge gained through Part 1 by outlining the practical steps required to design and run a successful OSCE, from preparation and planning through to implementation and post-OSCE considerations.

We have chosen to present Part II of this Guide as the evolving story of a fictional character Eva, an enthusiastic educationalist who is new to the concept of the OSCE. She is asked by the Dean of her institution to introduce the OSCE as a new form of assessment for the health care students graduating the following year. The knowledge and experiences she gains through this process are outlined in this Guide to assist others in implementing an OSCE for the first time or quality assuring their existing assessment processes.

Preparation and planning

Organisational structure

Large numbers of personnel are required in order to successfully implement an OSCE programme (Cusimano et al. 1994). Within higher education institutions there is usually a team

Practice points

- The assessment team would need to adopt new roles and responsibilities when setting up a new OSCE programme.
- A nominated OSCE lead needs to have an oversight of all aspects of the OSCE programme.
- An OSCE question bank needs to be developed and maintained in order to have a pool of quality assured and peer reviewed stations for use in various examination sittings.
- Examiner and Standardised Patient training are important elements of quality assurance and standardisation process.
- Post-hoc psychometrics provide valuable data for further quality assuring the OSCE questions and the programme as a whole.

responsible for overseeing assessment procedures. Changes to an assessment programme, such as the implementation of new methods of assessment should be undertaken with the help of the assessment team. It may be worthwhile for a small sub-committee to be formed from the members of the Assessment Team to lead on the introduction of the OSCE in the existing assessment programme. Following a successful implementation, ongoing review and quality assurance

Correspondence: Dr Kamran Khan, Department of Anaesthetic, Mafrq Hospital, Mafrq, Abu Dhabi, UAE; email: Kamran.Khan950@gmail.com

procedures can be continued by the Assessment Team, as for all other methods of assessment.

Within this sub-committee it may be beneficial to assign a single key person (the OSCE lead) with the overall responsibility and accountability for overseeing the development, organisation and administration of the examination (McCoy & Merrick 2001). This person should have expert knowledge or prior experience in conducting the OSCE. If this is not the case the chosen lead should gather information by reviewing the literature, attending workshops and seeking guidance from experts at other centres.

Eva's Story

One sunny morning as I was walking towards my office for work I met the Dean, who had just returned from the AMEE conference the previous week. After a quick greeting he invited me into his office. I wondered what was on his mind.

Over freshly brewed coffee he told me that he had been learning a lot about the OSCE at the conference. He thought it would be a good idea to introduce the OSCE to assess our students who would be graduating the following year and asked me to lead on this. I am always up for a challenge but this was a different one.

I did not know much about the OSCE, except that it is an assessment tool. I accepted the challenge but openly admitted that I would need to do a lot of homework, and would report back to him in due course. He was delighted that I had agreed to help and I looked forward to expanding my repertoire of assessment techniques.

I returned to my office and began to do some research on the topic; when, why and how was the OSCE developed? After this background reading, I was starting to grasp the theoretical principles of the OSCE but what I needed now was some practical advice on how to establish our own examination. Who could I ask?

I remembered my colleague and friend George, a Paediatrician at a neighbouring University. They had been using the OSCE for the assessment of medical students for some time.

The following week I had a chance to visit George to find out more. He showed me around their facilities and explained that I would need quite a bit of help with the workload. He advised that I initially consider forming a small organisational group, ideally including colleagues already involved with our assessment procedures. I decided to approach our pre-existing Assessment Team for support.

OSCE lead will have more time to address the academic considerations. Tasks such as the allocation of students to examination centres, distribution of examination paperwork and the handling of examination results should ideally be dealt by a dedicated administrative team.

Developing the larger team. Depending on the nature and format of the examination and the size of the institution there might be more than one site running the OSCE on the same day. At each site it may be helpful to develop a local organising team to oversee the practical aspects of the OSCE such as selecting local examiners and standardised patients, setting up the OSCE circuit and ensuring smooth running of the OSCE on the examination day. In smaller institutions members of the Assessment Team or their administrative support may perform such tasks.

Examination scheduling, rules and regulations

Setting the examination schedule. In any given academic year, there may be a need to schedule a number of OSCE sittings depending on the course curriculum requirements and the place of the OSCE within the broader assessment programme. It is common to run at least one OSCE for each year group of students per year. The exact timing of each examination should be primarily influenced by the institutional regulations and curriculum requirements, although venue and examiner availability should also be considered.

Setting an examination blueprint and examination length

Blueprinting and mapping. Blueprinting is the process of formally determining the content of any examination. In the case of an OSCE this involves choosing the spread of skills and the frequency with which each appears within an examination. Each blueprint for an OSCE should take into account the context of the examination, the content which needs to be assessed mapped to the curriculum and the need for triangulation, e.g. if any domains of assessment should be examined with the use of more than one assessment tool (Vargas et al. 2007). Part 1 of this Guide discusses the need to carefully match assessment methods to the skills, knowledge and attitudes being assessed, in more detail. In this way the OSCE should form only one part of the broader assessment programme.

The OSCE is primarily a competency-based assessment of performance in simulated environment (Khan & Ramachandran 2012), and therefore, principally assesses the skills-based learning outcomes. However, a detailed discussion of the learning domains which can be assessed by the OSCE is covered in Part 1 of the Guide. The blueprinting process should ensure that an appropriate sample of the skills-based curriculum is examined and it is mapped to the curriculum, i.e. the examination has adequate content validity. A blueprint normally consists of a two-dimensional matrix with one axis representing the generic competencies to be tested (e.g. history taking, communication skills, physical

Administrative support. Assessment of any kind inevitably creates a vast amount of administrative work. The OSCE is no exception to this and by ensuring there is adequate administrative support to meet these needs, the

Table 1. An example of an OSCE blueprint.

Topics	Procedural skills	Clinical examination skills	History taking	Total number of questions
Acute medicine	Q.1			1
ENT		Q.2 Q.3	Q.4 Q.5	4
Paediatrics	Q.6 Q.7			2
Geriatrics		Q.8 Q.9	Q.10	3
Total number of questions	3	4	3	10

examination, management planning, etc.) and the other axis representing the problems or conditions upon which the competencies will be demonstrated (Newble 2004). An example of a test blue print is shown in Table 1. Blueprinting can be done 'in-house' by the Assessment Team; however, in higher stakes examinations, a Delphi or other survey techniques may be used to agree on the topics to be included in the test blueprint. Questions can then be developed or chosen based upon the test blue-print.

Examination length (number of stations). In order to develop an examination blueprint, the examination length needs to be determined beforehand. This will depend on the number of stations within each OSCE and the length of each station. An OSCE station typically refers to one time-limited task given to the candidates generally lasting between 5 and 10 min. The reliability (reproducibility of the test results) and validity (the extent to which the test's content is representative of the actual skills learned or to be assessed) are both influenced by the number of stations and total length of the examination (Newble 2004). An appropriate and realistic time allocation for tasks at individual stations will improve the test validity. Whereas, increasing the breadth of the content, usually by ensuring an adequate number of stations per examination, improves reliability. In fact, the content specificity has been found to be a major contributor to poor reliability; hence competence testing across a large sample of cases is required before a reliable generalisation of candidates' performance can be made (Roberts et al. 2006).

The number of stations needed to generate a reliable score represented by either Cronbach's α or Generalisability (G) coefficient determines the examination length. A Cronbach's α or G value between 0.7 and 0.8 reflects an acceptable reliability for high stakes examinations. A detailed discussion on this topic is beyond the scope of this article and interested readers are advised to refer to AMEE guide 49 by Pell and colleagues (Pell et al. 2010) and AMEE guides 54 and 66 both by Tavakol & Dennick (2011, 2012). Work by Shavelson and Webb (2009) may also be of practical use. These concepts are also re-visited throughout the Guide.

For practical purposes decisions around test length generally need to balance reliability coefficients with feasibility and resource issues; but as a general recommendation, with well-constructed OSCE stations, an adequate reliability could be achieved with 14–18 stations each with 5–10 min duration (Epstein 2007).

Developing a bank of OSCE Stations

Before the stations are added to the bank they need to go through the processes of peer review and piloting. If available, psychometric data on individual stations might also provide useful information on station quality including its ability to discriminate between high-achieving and low-achieving candidates. These aspects are described in some detail below. A secure bank of robust and quality assured stations contributes significantly towards the better reliability and validity of the examination scores. The flow diagram presented here describes one approach to producing a bank of OSCE stations to meet the needs of the curriculum (Figure 1). It can be used as a step-by-step guide or adapted for individual requirements. A pre-existing bank of stations could also be updated or quality assured by following appropriate steps in the algorithm.

Choice of topics for new stations. In institutions where an OSCE is being set up for the very first time, the examination blueprint governed by the curriculum outcomes would act as a good starting point to identify the topics for writing new OSCE stations. In places where an OSCE station bank already exists, the OSCE lead or subject experts could review this to identify gaps in the assessment of certain skills or domains. The need for new stations could also arise if the curriculum is modified or learning objectives of the modules are changed. The assessment of competencies should always be aligned to the teaching and learning that has taken place as specified by the course curriculum. Occasionally assessment content is influenced by the regulatory authorities as in the case of General Medical Council (GMC) in the UK, which stipulates that medical graduates should be able to demonstrate certain competencies (GMC 2009).

Once the areas for assessment have been identified it is important to ensure that the clinical skills which are expected to be performed by the candidates can be realistically assessed using an OSCE format and in the limited time allocated for each station (Vargas et al. 2007). Part 1 of this Guide describes in some detail what the OSCE can assess most appropriately.

Choice of station writers. The OSCE lead has the responsibility of identifying appropriate people to design and write the OSCE stations. If a pool of trained examiners already exists, it would be an obvious choice to seek volunteers for question writing from this pool. Otherwise subject experts can be asked

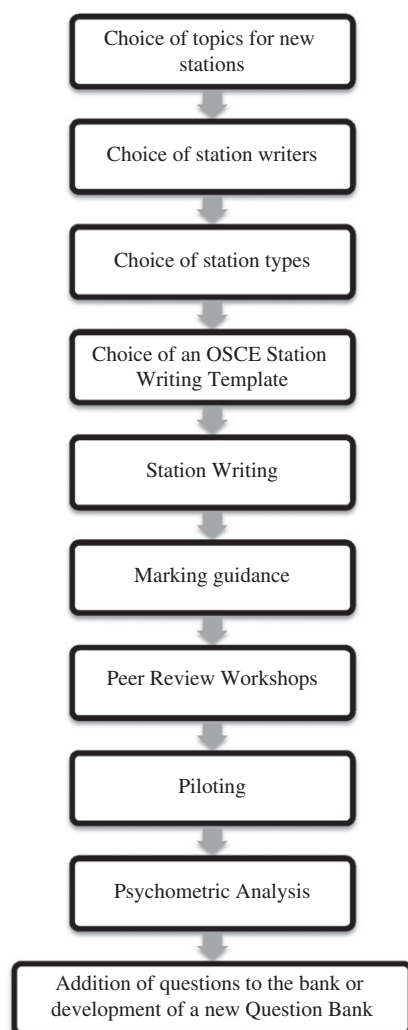


Figure 1. Flow diagram for the development of an OSCE question bank.

to help with writing. It is essential for station writers to be familiar with the underlying principles of the OSCE in order for them to produce appropriate work. Brief orientation sessions or written instructions could be developed for people new to this task.

Choice of station types. The OSCE lead or the person coordinating the station writing should advise the question writers about the type of new stations needed. An understanding of the different types of OSCE station formats is essential in the choice of appropriate station types for various assessment outcomes (Table 2).

The choice of OSCE station writing template. Once the type of station has been chosen, an appropriate template for station writing should be developed or used. A template helps authors to develop stations in a similar format to others within the bank. Such a standardisation prevents disadvantaging the candidates by posing questions in unfamiliar formats and helps to maintain the reliability of the scores. We have shown an example of an observed OSCE station writing template. This template is supplemented with an example of a station

designed to assess the focussed history taking and respiratory examination of a patient with asthma (Appendix 1).

Station writing. The different sections of the template highlight the information that should be considered in order to write a successful OSCE station. Each of these sections is shown below with an explanation of the type of information required (Table 3).

Marking guidance. The marking guidance for each station would depend on the scoring rubric chosen as a standard for all OSCE examinations. There is a detailed discussion on the different types of scoring rubrics later in this Guide. If it is a checklist or a rating scale the author should develop it as they write the station. Such a checklist should reflect the outcomes being assessed. The same is applicable for rating scales if these are specific to the stations. If a global rating scale is to be used in isolation then the marking criteria for individual stations may not be required. AMEE guides 49, 54 and 66 discuss further the impact different scoring rubrics can have on OSCE outcomes (Pell et al. 2010; Tavakol & Dennick 2011; Tavakol & Dennick; 2012).

Peer review workshops. Running review workshops with examiners is one way of quality-assuring new OSCE stations. Once the examiners have written the new stations, they are invited to bring these to the workshops where delegates can review stations written by others, often in small groups. The presence of the authors for individual stations at the workshops ensures changes and clarifications are made more easily.

In addition to looking at the clinical accuracy and appropriateness of the tasks involved in the station the peer review process can help to identify validity issues as well. A simple questionnaire could be used for this purpose, an example of such a questionnaire is shown in Appendix 2.

Piloting. After the peer review process by the examiners, piloting of the stations helps to identify any issues with the practicality and allocation of time for the tasks. If required, changes can then be made to the stations to improve their quality (Whelan 1999). Initial psychometric analysis on reliability and station quality could also be done at this stage. In the case of any problems with a particular station it should be redesigned and then re-piloted. Piloting often takes place during mock or low-stakes examinations which may have the additional benefits of orientating candidates to the OSCE and providing them with immediate feedback on their performance. If individual stations are piloted within the circuit of a high stakes examination it is essential to inform the candidates about the inclusion of a pilot station and that its scores will not influence the overall examination results. In order to get valid and reliable data on the pilot stations included in real examinations, the identity of such stations is not disclosed beforehand.

Psychometric analysis. We have briefly discussed relevant aspects of psychometrics in the section on Examination Length earlier. With respect to development of new stations,

Table 2. Types of OSCE stations.

Station type	Description	Examples	Benefits	Limitations
Observed station	An examiner is present throughout the duration of testing	Communication skills Procedural skills Clinical examination	Direct observation allows assessment of the higher levels of the learning domains. Immediate feedback can be given for formative assessment.	Resource intensive with one examiner needed per station
Unobserved station	No examiner present throughout the testing period. Answers may be submitted on paper either after each station or following the completion of the examination	Interpretation of clinical information e.g. X-rays, pathology specimens, blood results. Prescribing skills. Information Technology Skills.	Examiners not required on such stations.	No direct observation of performance. OSCE station may not be necessary and alternative assessment tools may be as effective at assessing the cognitive skills in question.
Technology enhanced station	A station involving the use of technological advances such as part-task trainers or high-fidelity manikins to assess skills that would otherwise be difficult to assess in the OSCE format.	Intimate clinical examinations with use of part-task trainers, e.g. rectal examination. Complex decision making skills and the management of acutely unwell patients with the use of high-fidelity manikins.	Increases the scope of potential OSCE stations, allowing assessment of learning domains which could not be assessed effectively using traditional OSCE stations.	Personnel must be trained in the use of the equipment. Equipment failure. Initial cost of new equipment and maintenance costs.
Linked stations	Two consecutive stations are based upon the same clinical scenario or information. These may be observed or unobserved.	Observed examination of the respiratory system in the first of two stations. Unobserved documentation of findings and management plan in the second station.	Greater number of skills can be assessed per scenario Efficient use of examiner resources.	Needs careful circuit planning, as candidates can neither start on the second of a pair of linked stations nor end the examination on the first.

if a complete set of new questions is used in a mock OSCE then the psychometric analysis will indicate the overall reliability of the set of questions. Use of G theory will indicate the number of similar stations needed to achieve a good reliability by performing D or decision studies on the data (Shavelson & Webb 2009). Application of Item Response Theory (Downing 2003) will also be able to yield data highlighting the sources of variability or error. This theory can be used if one or more stations are piloted in real examinations. A detailed discussion on this topic is again beyond the scope of this Guide, but it is discussed in some detail in AMEE Guide No. 49 (Pell et al. 2010), and G theory is comprehensively covered in AMEE Guide No. 68 (Bloch & Norman 2012).

Choosing a scoring rubric and standard setting

Stevens and Levi (2005) have defined a scoring rubric as 'an assessment tool that delineates the expectations for a task or an assignment'. Various scoring rubrics are used to mark different types of assessment. There are two main types of scoring rubrics, analytical and holistic.

Analytical scoring (checklist scale). A checklist is a list of statements describing the actions expected of the candidates at the station. It is prepared in advance, following consultation with the team designing the OSCE stations and in line with the content and outcomes being assessed. Checklists could be 'binary', yes/no (performed/not performed), i.e. candidates

are marked based on whether or not an action was performed, without any discrimination for the quality of the actions. Such checklists are may not be able to discriminate between lower and higher levels of performance. Alternatively, checklists can have 5–7 point rating scale, which allows the examiners to mark candidates based upon the quality of the actions. Such checklists with rating scales are different from global ratings (holistic scoring), which are described later.

Traditionally, a key strength of binary checklists has been their perceived ability to provide an objective assessment and thought to lead to greater *inter-rater reliability*. In fact, such checklists were originally used by Harden when he first developed OSCE techniques as shown in the first part of this Guide (Harden et al. 1975). There is, however a growing body of evidence which has called this view into question, showing that objectivity does not necessarily translate into greater reliability (Wilkinson et al. 2003). This is particularly applicable if expert examiners are used in an OSCE (Hodges & McIlroy 2003). An example of binary checklists and rating scales (which can be seen as mini global ratings as they rate one element of the overall consultation), is shown in Table 4.

Holistic scoring (global rating scale). Compared with checklists, which are task specific (Reznick et al. 1998), global rating scales allow the assessor to rate the whole process. Consider the performance of an expert who may not follow a pre-determined sequence of steps as outlined by a checklist, yet still performs the task to a high standard with fluidity and ease. In this situation an overall (global) assessment of the

Table 3. Guidance for completing the question writing template.

Template section	Information required
Station information	
Subject/Topic	Chosen as described at the beginning of this section.
Level of candidate	The competencies that need to be demonstrated should be appropriate for the level of training of the candidates.
Competencies to be assessed	As described in part one of this Guide.
Station duration	This should generally be standard for all stations as determined by the OSCE committee as described in the examination length section.
Information for the site organisers	
Standardised patient (SP) age and sex	This information is key for the organising team to find appropriate SPs for the stations.
Resources and equipment needed	The organising teams need this information to equip the station and to maintain standardisation across sites if an examination is taking place at multiple sites.
Setting up the station	This information is needed to ensure that the position of the chairs, tables and couches do not interfere with the tasks in questions. For example candidates should be able to approach the SP for examination from the right side and the examiners are able to view the whole process from where they are seated.
Instructions for candidates	
What is the scenario (briefly)?	This is the key information for the candidates about the station.
Who and where they (Candidates) are?	It is essential for the candidates to know in what capacity they are having a consultation with the patient, so that they can provide appropriate information to the patients and the examiners.
What has already happened	If the scenario is setup in such a way that a nurse has seen the patient and some information about the blood pressure and temperature etc. could be available to the candidates if they asked for it, then it should be made clear at this stage.
What are the candidates expected to do	These are the skills to be demonstrated by the candidates at a given station, e.g. take a focused history and examine the respiratory system
What are the candidates not expected to do?	If due to limitation of time candidates are not expected to perform General Physical Examination as a part of the respiratory system examination then it should be clarified here.
Supplemental data	If any investigations could be provided to the candidates then it should be included in this section. For instance, 'if you require the results of any relevant blood tests you can ask the examiners
Template section	Information required
Information for the examiner	
Brief background to the scenario	This section is similar to 'What is the scenario?' section above.
Examiner's role	It is very important to make clear here, in what way the examiners are allowed to interact with the candidates if at all. Are they supposed to observe only or ask some questions as well?
What are the objectives of the station	Again similar to 'What are the candidates expected to do?' section above.
What information they might be able to provide the candidate?	The examiners, if required, can provide information such as reports of blood tests or X-rays.
What information they should not provide the candidate?	The examiner should not volunteer any information candidates are expected to explore for themselves.
Clinical information relevant to the station	If examiners from disciplines other than the one in question are used for assessment then this information should be able to enable them to mark the candidates fairly. This becomes even more important, especially if generic marking schemes are to be used for assessment as described later in this Guide.
Simulated patient information (All the information in this section is essential in maintaining the standardisation of the station, especially if the examination is being conducted at split sites)	
Who are they?	As set out in the candidates' instructions earlier.
Their social/economic background (if applicable)	This needs to be brief, considering the station length and time allowed for the candidates to complete this task.
Presenting complaint and past history	Should be succinct and focused for the purposes of the station.
Details of current health problems and medications	Again this information should be relevant and to the point.
Details of their concerns/perceptions	If a station involves dealing with SP concerns then it is important to standardise this aspect.
What they should say (their agenda) and what they should not say	In certain stations this information would not be needed, for instance where the candidates are expected to examine only.
What they should ask (questions)	The SPs should be allowed to ask standard questions when talking to the candidates.
Specific standardisation issues (specific answers to specific questions, please stay with the script)	This information could be vital if the candidates are expected to ask some particular questions.

performance is required in order to accurately reflect the skill of the candidate. Global scales allow examiners to determine not only whether an action was performed, but also how well it was performed. This tool is therefore better for assessing skills where the quality with which it is performed is as important as performing it at all. An example might be the assessment of a candidate's ability to empathise with patients in communication skills stations. Hence holistic scales are more useful for assessing areas such as judgement, empathy,

organisation of knowledge and technical skills (Morgan et al. 2001; Hodges & McIlroy 2003). Global ratings differ from checklist rating scales described above by the virtue that global ratings take a more holistic view of the overall performance at a station compared to a rating scales looking at one aspect alone.

Global ratings are being increasingly used over checklists for marking at OSCE stations, as there is now evidence to suggest that they show greater inter-station reliability, better

Table 4. Comparison of binary checklist and rating scale.

Station on history and examination of an asthmatic patient: two possible types of checklist for scoring the examination of the chest	
Binary check list	Checklist using rating scales
Candidate performs an examination of the chest 1. Introduction Yes ☐ No ☐ 2. Obtaining consent Yes ☐ No ☐ 3. Appropriate exposure Yes ☐ No ☐ 4. Professional approach Yes ☐ No ☐ 5. General physical examination Yes ☐ No ☐ 6. Inspection Yes ☐ No ☐ 7. Palpation Yes ☐ No ☐ 8. Percussion Yes ☐ No ☐ 9. Auscultation Yes ☐ No ☐ 10. Advising patient at the end to cover up the exposed areas and thank for cooperation. Yes ☐ No ☐ Score awarded X/10 Add the scores and award out of 10	Candidate performs an examination of the chest 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 1. Unstructured approach 2. Structured approach but completes less than 50% of key steps. 3. Structured approach but completes more than 50% of key steps. 4. Structured approach and completes a majority of key steps. 5. Structured approach and completes all key steps. Such a check list should then be accompanied by a list of key steps. Uses alcohol rub before and after examination and, when appropriate uses gloves Seeks permission to examine, and explains the nature of examination Offers/Asks for chaperone where appropriate Asks the patient if any areas to be palpated or moved are painful Positions the patient correctly and comfortably, then uses a methodical, fluent and correct technique Does not distress, embarrass or hurt the patient unduly Examines, or suggests examining, all the relevant areas Completes the task, covers up the exposed areas and thanks the patient

construct validity, and better concurrent validity compared to checklists (Turner & Dankoski 2008). Further information on the impact of scoring rubrics can be found in AMEE Guides 49, 54 and 66 (Pell et al. 2010; Tavakol & Dennick 2011; Tavakol & Dennick 2012).

Standard setting. Standard setting refers to defining the score at which a candidate will pass or fail. There are a number of methods that can be used for this purpose. In the Norm referenced methods the scores have meaning to each other and the pass/fail scores are determined by the relative scores of candidates, e.g. the Cohen method (Cohen-Schotanus & van der Vleuten 2010). In a norm referenced examination the standard set is based upon peer performance and can vary from cohort to cohort. It is, therefore, possible that in a 'poor' cohort, a candidate may pass an examination that they would have otherwise failed if they took the examination with a 'stronger' cohort. For this reason norm referencing is usually deemed unacceptable for clinical competency licensing tests, which aim to ensure that candidates are safe to practice. In this case a clear standard needs to be defined, below which a doctor would not be judged fit to practice. Such standards are set by Criterion referencing in which scores have absolute meanings to the domains of assessment. Angoff (1971) and Ebel (1972) are two commonly used methods for this purpose. The criterion methods of standard setting are performed before the examination by a group of experts who look at each test item to determine its difficulty and relevance. Although both Angoff and Ebel methods are well established, these were initially developed for tests of knowledge such as multiple-choice examinations and it may not always be appropriate to extrapolate these methods to tests of

performance, i.e. the OSCE (PMETB 2007). Other absolute methods which may be of relevance include Borderline Group and Contrasting Groups Methods; readers can find more about these in the articles written by Kaufman (2000) and Kramer (2003). A detailed discussion of each of these methods and their pros and cons is beyond the scope of this Guide and interested readers are also referred to AMEE Guide No. 18 on Standard Setting in Assessment by Friedman (Friedman Ben-David 2000) and articles by Tavakol (2012) and Pell (2010).

Developing a pool of trained examiners

In maintaining the reliability of the scores in an OSCE examination, consistent marking by trained examiners plays a pivotal role. Examiner training is an ongoing process whereby new examiners are added to the pool and the existing examiners are provided with refresher training. This section deals with the process of examiner training and retention.

Identification of potential examiners. The reliability of the scores generated by the examiners not only depends upon the consistent marking by the examiners but also their clinical experience relevant to the OSCE station. It is common for doctors to assess doctors and nurses to assess nurses, however, skill-matching can add a degree of flexibility and has resource and financial implications. In order to lessen the burden of finding adequate numbers of doctors to act as examiners; there have been instances where non-physician examiners have been used in medical examinations. There is literature suggesting that simulated patient scores have a good correlation with physician scores (Mann et al. 1990;

Box 1. Outcomes for examiner training.

Learning outcomes for examiner training sessions

- To understand the scope and principles of the OSCE examination
- To maintain consistent professional conduct within the examination
- To understand and use the scoring rubric in order to maintain standardisation
- To provide written feedback on performance if required in summative examinations
- To provide verbal feedback at the end of station in the formative examinations
- To ensure confidentiality of the candidates' marking sheets
- To understand the procedures for inappropriate or dangerous behaviour by candidates

Cohen et al. 1990). However one study suggests the agreement between physician and non-physician scores, although good for check-list scoring, does not extend to global scoring (Humphrey-Murto et al. 2005a). Whether expert or non-expert examiners are chosen, training all examiners will reduce examiner variation (Newble et al. 1980; van der Vleuten et al. 1989) and improve consistency in behaviour, which may improve exam reliability (Tan & Azila 2007).

Most physician examiners are sourced from local hospitals or community practices and it is helpful to approach those with a prior interest in medical education. In most professions the examiners are not financially remunerated as this is seen as a part of their responsibility towards teaching and training; however those who do volunteer describe an enhanced sense of duty and an insight into learners' skills (Humphrey-Murto et al. 2005b).

Examiner training workshops. Examiner training sessions should ideally take place well in advance of the examinations. The level of training will depend upon the background and ability of the examiners (Newble et al. 1980; van der Vleuten et al. 1989). As with any other teaching and learning activity the outcomes of the examiner training workshops should be explicit (Box 1).

These sessions can be organised in any format but generally include group discussions about some of the above topics, followed by the opportunity for the examiners to mark Mock OSCE or videos of real OSCE.

Although examiners tend to maintain and further develop their skills by regularly assessing, the need for refresher training can be driven by a change in the format of examination or scoring and also by changes in the requirements of the institutions or regulatory bodies. Such refresher training could be delivered using online resources or by further small group sessions.

Developing a pool of trained standardised patients

Patients form an integral part of an OSCE with many of the stations requiring active patient participation. Collins & Harden (1998) refer to the continuum of patients used in clinical examinations, from the real patient with clinical signs who receives no training to the rigorously trained simulated patient. The recruitment and training of each type of patient will differ depending upon their role within the examination.

Although the terms 'simulated patients' and 'standardised patients' are used interchangeably, a simulated patient is a usually a lay person who is trained to portray a patient with a specific condition in a realistic, and so standardised way (Cleland et al. 2009). Standardised Patient (SP) is an umbrella term for both a simulated patient and an actual patient trained to present their condition in a standardised way (Barrows 1993). Standardisation in the term 'standardised patient' relates to the consistent content of verbal and behavioural responses by the patient to stimuli provided by a candidate (Adamo 2003).

Recruitment of standardised patients. The type of patient required for each OSCE station will depend upon the desired outcomes of the station and the role expected to be played by them. If the station requires the candidate to elicit a specific clinical sign, e.g. a heart murmur, a real patient with the murmur in question must be used. However, if the focus of the station is to determine if the candidate can competently examine the cardiovascular system (regardless of any clinical abnormality) a 'healthy' volunteer can be used instead. Certain stations, such as history taking and communication skills stations will generally require the use of trained simulated patients. AMEE Guides Nos. 13 and 42 provide a detailed discussion on choosing the correct 'patient type' for the examination in question (Collins & Harden 1998; Cleland et al. 2009).

Patients can be recruited in a number of ways; real patients with clinical signs can be accessed through contacts with primary and secondary care physicians. A doctor previously known to the patient and responsible for their care may be the most appropriate person to make initial contact (Collins & Harden, 1998). Recruiting patients with common conditions that remain stable over time is easier than finding patients with rare and unstable disease and this should be taken into account at the time of blueprinting and station development. Healthy volunteers can be found through advertising in the local press, contacts with local educational institutions and by the word of mouth. Actors are commonly used for complex communication issues such as breaking bad news and for high-stakes examinations (Cleland et al. 2009). Highly trained professional actors are likely to incur significantly higher costs than volunteers and real patients who may be remunerated by the reimbursement of expenses alone.

In many large institutions a standardised/simulated patient co-ordinator is employed to undertake the selection process keeping in mind the ability, suitability and credibility of the SPs. Each of these areas is discussed in detail in AMEE Guide 42 and is beyond the scope of this guide (Cleland et al. 2009).

Standardised patient training. All standardised patients will require training, but real patients and simulated patients (actors) will require different levels of input. All will need to understand the importance of portraying the clinical conditions in question, reliably and repeatedly and the need for standardisation between candidates. In some cases the pre-examination briefing on the day may be adequate for this purpose; generally simulated patients for role play in more

complex scenarios will require dedicated training in advance of the examination.

In addition to their use in the OSCE, simulated patients are often used for teaching skills to the medical students outside the examination settings. It may be convenient and cost effective to train groups of simulated patients together to be used for a variety of purposes within the institution. In depth discussions on simulated patient training workshops can be found in AMEE Guides Nos. 13 (Collins & Harden 1998) and 42 (Cleland et al. 2009) and are not reproduced here. In addition there are associations dedicated to educating standardised patients such as the Association of Standardised Patient Educators, who provide leadership, education and structure to the training and assessment of standardised patients (Turner & Dankoski 2008). Although there is no real consensus in the literature as to the sufficient duration of training for each simulated patient, one estimate suggests it may take up to 15 h to adequately train a simulated patient dependent on the role, experience and adaptability of the person (Shumway & Harden 2003).

Once training is completed each standardised patient's performance needs to be quality assured before being used in a high stakes examination. Simulated patients may be videotaped and their performance evaluated by an independent group of trainers (Williams 2004). Alternatively new simulated patients could be used for the first time in mock OSCEs and feedback from candidates and examiners could be used to quality assure their performance (Stillman 1993).

If there is a standardised patient co-ordinator, they should hold a bank of trained patients who can be called upon for subsequent examinations. Ideally individuals within this bank should be trained to perform multiple roles, this will increase the flexibility and maximise the potential to find the right person for the right scenario (Whelan 1999).

Standardised patients are a valuable resource, it is important to keep them interested in the role by using them regularly, remunerating appropriately and always expressing thanks for their input (Cleland et al. 2009).

Running the OSCE

Administrative tasks

As previously described, any form of examination generates considerable administrative work. We describe here the key administrative activities that may need consideration in order to ensure the smooth running of an OSCE (Box 2).

All relevant information pertaining to the implementation of the OSCE could be held within a procedure manual for future reference. This may include lists of trained examiners, trained SPs, sources of equipment and catering facilities.

Choosing an OSCE venue

The OSCE venue should be booked well in advance bearing in mind the number of stations and candidates. In addition to housing the examination itself, the venue should ideally have the capacity for briefing rooms, administrative offices, waiting rooms for patients and examiners, quarantine facilities and

refreshment areas. Stations may be accommodated in several small rooms similar to outpatient clinics or alternatively a larger room can be turned into 'station areas' with the use of dividing screens. Individual rooms have the advantage of increased confidentiality and low noise levels but may make the signposting of the circuit more challenging. Some institutions have a special site allocated specifically for the examinations.

Eva's Story (continued)

I have learnt so much about OSCEs in the past year! I had no idea so much preparation was going to be required in introducing this new examination. As the OSCE lead I was overseeing all of the academic considerations that I have just described to you.

It has taken some time but we now have a bank of questions designed to assess the final year medical students, these are blue-printed and mapped against the curriculum and have been quality assured at a peer-review workshop.

We have spent the last few months identifying and training our new OSCE examiners. George was a real help here, as he came along to describe to them how the OSCE worked at his University. Most of the new examiners were supportive of this change to assessment although there were a few who were quite resistant. Having the background knowledge of the advantages and disadvantages of the OSCE was really helpful in the debate that ensued.

We have managed to find some volunteers to act as patients in the OSCE and have identified some real patients with good clinical signs for our clinical examination stations.

We have decided to run a pilot examination with a group of the current interns about a month before the real examination of the final year students. In this way we can check the practicalities of the stations and ensure the examiners and patients are comfortable with their roles. There are still so many things to think about to ensure the smooth running on the big day. I have made a list of all the essential requirements for running an OSCE and share it with you now.

Setting up the OSCE circuit and equipment

The OSCE circuit. The circuit is the term used to describe the setup of stations for the seamless flow of candidates through the examination. Each candidate will individually visit every station within the circuit throughout the course of the examination. The number of candidates in each sitting should, therefore, be equal to the number of stations, unless rest stations are used as described below. Each candidate will be allocated a start station and move from station to station in the direction of the circuit until all stations have been

Box 2. Common administrative tasks for OSCE.

Common administrative tasks for the OSCE

Allocation of students to examination centres

- If examinations are to be held at multiple sites, planning is required to ensure that wherever possible examiners do not know the candidates and any candidates with disabilities are sent to centres with appropriate facilities.

Transport and reporting instructions

- Candidates must be provided with comprehensive instructions about where to report at the examination centre. In some circumstances transport may need to be arranged for large groups of candidates.

Distribution of paperwork

- Station information, candidates' lists and mark sheets need to be printed, collated and distributed to all examination sites. Mark sheets should be pre-populated with candidates' details to minimise time required during the examination.

Selection of standardised patients

- Once equipped with the station information it is necessary to identify appropriate SPs from the trained pool for all stations. Commonly, more than one SP for each station is identified, as fatigue may occur if the station is to be run several times in the day. In addition it is also advisable to invite a number of reserves. They should receive their scripts and reporting instructions in advance.

Selection of examiners

- Once the station information is known appropriate examiners must be selected from the trained pool, taking into consideration the decisions made regarding expert versus non-expert examiners. Reserve examiners should always be invited.

completed. The local organising team will usually be responsible for setting up the circuit.

Circuit with rest stations. The addition of rest stations allows a break for the candidates and examiners and may allow for the addition of an extra candidate if required (Humphris & Kaney 2001). Care should be taken to keep this station private, so that the candidate at this station cannot over hear what is being said at the other stations. It should be clearly marked, and candidates should be informed of its presence before the start of the examination, although ideally students will have had practice sessions to familiarise themselves with the examination circuit. It is extremely important to bear in mind that the circuit cannot start or finish with a rest station. If the rest stations are at the beginning or the end of a circuit a candidate will end up missing one or more stations. For this purpose the rest stations should be interspersed within the live stations.

Considerations for individual stations. In setting up individual stations, care must be taken to allocate space appropriate to the tasks, equipment and the personnel. For example, an unmanned station containing investigation results and some written questions would need just enough room for a table and chair, whereas a resuscitation station would need enough space for a manikin, a defibrillator and an examiner. The stations should provide an appropriate environment for the candidates to perform the procedures. For instance, adjustable lighting for Fundoscopy or a quiet area for auscultation of the chest should be provided as appropriate (McCoy & Merrick 2001). Some stations may also require power sockets for the equipment.

The equipment. The equipment required for each OSCE station is included in the documentation developed at the station writing stage. All equipment should be sourced well in advance of the OSCE, and checked to ensure that it is in good working order. There should be spare equipment and batteries available on the day in case of breakages or breakdowns.

Box 3. Examination day briefings.

Information for examination day briefings

Candidates briefing

- A description of the circuit including their start stations, rest stations and pilot stations.
- Reminders of rules and regulations
- Quarantine procedures
- Fire procedures

Examiners briefing

- The objective of the examination, e.g. formative/summative
- To check students' identity at the start of the station
- An overview of the scoring rubric and how to complete the mark sheets
- The importance of keeping stations and candidates' scores confidential
- Not to talk to the students any more than what is allowed in the script
- To treat all candidates equally
- The procedures for reporting concerns about candidates
- Completing feedback after the examination
- Fire and quarantine procedures

Standardised patient briefing

- The importance of standardisation between candidates
- Their role in providing feedback
- Rest-breaks and refreshment facilities
- Fire procedures

Decisions ought to be made about candidates' use of their own equipment during the examination. If candidates are expected to bring their own stethoscopes for instance, they should be informed of this.

If more advanced equipment is required such as high fidelity human patient simulators there must be personnel available who are able to programme and run these, as most examiners will not be familiar with such equipment.

Examination day briefings. On the day of the examination there should be separate briefing sessions for the candidates, examiners and SPs. If there has been prior training and if written instructions have been provided they need only be brief and should be kept succinct. Key information that may be included is outlined below (Box 3).

Table 5. Common problems and troubleshooting tips.

Problem	Potential solution
Variable performances by SPs affecting station standardisation	Occasionally SPs may change their behaviour between candidates or provide unsolicited information. Robust selection and training procedures should minimise these issues. Examiners should also be aware of this potential problem and be willing to intervene between candidates if necessary.
Equipment failure	There should always be spare equipment readily available at hand. If candidates lose a lot of time waiting for spare equipment it may be possible for them to retake the station at the end of the examination.
Unpredictable behaviour of candidates	Nervous candidates under stress can often act in unpredictable ways. In particular, getting lost in the venue or on the OSCE circuit. Adequate support staff should be available to help direct candidates and answer any queries. Examiners may have to prompt candidates to move on at the correct time if bells or voice commands are missed.
Removal of instructions or equipment from stations by candidates	Instructions can be firmly secured to a table or lectern. Examiners and support staff should be vigilant for candidates leaving stations with equipment
Removal of mark sheets or station information by examiners	This may preclude the station form being used in subsequent sittings and examiners should be warned that no documentation must leave the station. Support staff collecting documentation prior to examiners leaving the station can reduce the chance of this occurring.

Running the OSCE circuit and troubleshooting

Running the circuit. The movement of the candidates from one station to another can either be managed by ringing a bell manually or by using automated PowerPoint™ presentations set up with voice commands clearly instructing the candidates and the examiners. The OSCE starts with the command 'Start Preparation', during which time the candidates read the question, followed one minute later with instructions to 'enter the station'. The next instruction could be 'one minute left' and the station would end a minute later with the command 'move on'. During a formative examination an additional command 'start feedback' at an appropriate time interval could also be included. The cycle is repeated for the duration of the examination. This system may be preferable to the use of bells as it reduces the confusion as to what each ring of the bell signifies. However, if an automated system of commands is used, a back-up in case of technical failure is essential, which could be a simple stopwatch and a bell.

Once the examination is started there should be personnel available to ensure that the candidates move in the right direction. If any SPs need a break they should be replaced promptly with reserve SPs as described earlier. At the end of the examination the marking sheets are collected and the stations are reset for the following run of the circuit if needed.

Quarantine. Quarantine refers to separating those candidates who have completed the examination from those who have yet to take it on the same day. The same set of OSCE stations may be in use for both morning and afternoon sessions, allowing exchange of information if the morning candidates are allowed to leave prior to all the afternoon candidates arriving. This may lead to a perceived unfair advantage to the second set of candidates. To resolve this issue, candidates scheduled for the early circuits should be 'quarantined' in a separate room until all of the later candidates have arrived and registered. Mobile phones and other devices with the means for remote communication should not be permitted in the examination centres.

Trouble shooting. On the day of the examination a number of issues can arise, some common issues and their potential solutions are described below (some of this information is

taken from the Queens University of Belfast's website on OSCE training available at http://www.med.qub.ac.uk/osce/background_Dilemma.html) (Table 5).

Eva's Story (continued II)

So we did it! What an exhausting day it was but we are very pleased with how it went. The candidates, patients and examiners all turned up and knew what to do. The meetings and planning we had been through were all worth it. The examination ran smoothly, except for a few hiccups with equipment failure, but we had anticipated it and were able to replace faulty tools with spares. The patients also became quite exhausted by the end of the day and I think we will recruit more reserves in future.

Now that we've got the OSCE itself out of the way, we can all let out a huge sigh of relief but there is still quite a lot of work to do. I was reminded of this as soon as I arrived in my office today to check my emails; there were a few from students asking when they would get their results. We need to collate all the marks and publish them. I am looking forward to analysing the candidates' results and the psychometrics of our OSCE; we should be able to extract some really valuable statistics to help us in improving things for next time. I've been in discussions with our psychometrician who has already helped a great deal and will now be invaluable.

The quality assurance process has been important to us from day one; we didn't just want to put on OSCE for the sake of it, we wanted a reliable and valid assessment that assessed skills not tested by our other tools. This process continues now, with feedback, evaluation and psychometrics. We can use all of this information to keep improving our OSCE for future students.

Post-OSCE considerations

Handling results

Following the examination the mark sheets are collected and cross-checked for accuracy and any missing scores. The examiners are contacted if any corrections need their

verification. These results are put in appropriate spreadsheets and cross-checked again in preparation for the examination boards for ratification as described below.

The examination boards and ratification

After compilation the results are made available to the examination boards for the purposes of ratification. The examination board ratifies the results and signs them off as accurate. In case of any doubts the results are verified again. In cases of poor performance or failure the penalties are decided at these meeting and later conveyed to the students.

Publication of results

After the ratification the publication of accurate results is the final responsibility of the Assessment Team. The results could be made available online as well as sent to the students as hardcopies.

Complaints and appeals

There may be mitigating circumstance appeals or complaints made by candidates or examiners that need to be dealt with fairly and promptly after each examination. There will often be institutional policies and procedures to follow under these circumstances. Valid complaints may help to inform changes to the examination as a part of the quality assurance process.

Quality assurance

The quality assurance of each examination is a continuous process repeated with each examination cycle. Although many quality assurance procedures take place following the OSCE, quality assurance by training examiners, peer reviewing stations and ensuring standardisation are also quality assurance measures that take place before the conduct of the examination. The Figure 2 highlights the factors contributing to quality assurance. Those that have not yet been addressed will be described in more detail (Figure 2).

External examiners

External examiners may be invited from different institutions to inform and comment on whether academic standards are being maintained between institutions and also to ensure the assessment process measures student achievement rigorously and fairly and is conducted in line with policies and regulations.

Post-hoc psychometrics

Post-hoc analysis of OSCE results allows the determination of the reliability of scores generated by the examination. This topic has been briefly dealt with in the sections on Station Length and Station Bank development. Although a detailed discussion is beyond the scope of this Guide, we would like to further address the concept of reliability at this stage for the sake of completeness.

The reliability of OSCE scores can be measured as Cronbach's α or G coefficient as mentioned earlier. Each of



Figure 2. Elements of OSCE quality assurance.

these coefficients represents the error in the scores generated by the OSCE. A coefficient of 1 means there is no error in the scores and all variance is true variance. A Cronbach's α or G coefficient of 0.7 to 0.8 is taken as an acceptable level of reliability for high stakes examinations. PMETB in the UK advocates a minimum reliability coefficient of 0.9 as a standard for high stakes Royal College examinations (PMETB 2007).

Application of Cronbach's α allows the detection of the OSCE stations which are main sources of error, by removing one station at a time from the analysis and looking at the reliability of the remainder. Application of G theory allows the identification of various other sources of error including the items, assessors and interaction of candidates with items and assessors etc. Item Response Theory also generates results somewhat similar to the G theory, but does not have the capacity to predict the reliability if the number of the stations was altered.

It is essential to perform psychometric analysis on the OSCE results and use the outcomes to enhance the quality of the examinations. Departments and institutions running OSCEs should seek help from Psychometricians in this respect. AMEE Guides 54 and 66 address post-hoc psychometrics (Tavakol & Dennick 2011, 2012).

Evaluation

The feedback on the examination process provided by the examiners can be used to improve the quality of the stations and organisation of the future examinations. Generally after each sitting of the OSCEs the examiners are invited to provide written comments on the individual stations they had examined on (Kowlowitz et al. 1991). Any issues such as undue difficulty of tasks, lack of clarity of instructions for the candidates and appropriateness of tasks for completion in the

allocated time are highlighted and addressed based on this information.

Candidates may also be invited to provide feedback on their experience of the examination as part of the quality assurance process (Williams 2004).

Conclusion

Part I of this Guide introduced the concept of the OSCE and explained the theoretical principles underlying its use as one part of a battery of assessment tools. Part II has focussed more on the organisational and practical factors for consideration while setting up an OSCE.

The key strength of OSCE is in its standardisation and reliability when compared to older forms of performance assessment, this reliability must not be compromised by poor planning or insufficient training of station-writers, station-examiners or standardised patients.

Organising and planning an OSCE from scratch is a huge task which requires a lot of logistical groundwork and training for all those involved. Good management and awareness of potential problems make the actual running of the OSCE easier. The quality assurance processes include post-hoc psychometrics to determine reliability and stations' quality. Together with the evaluation this psychometric data helps to improve the future examinations.

The instructions and advice in this two part Guide should help planners and faculty through every stage of the organisation of an OSCE, from understanding and application of the underlying theory to the administration, organisation, evaluation and quality assurance.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

DR KAMRAN KHAN, MBBS, FRCA, FAcadMed, MScMedEd, at the time of production of this manuscript was a Senior Teaching Fellow Manchester Medical School and Consultant Anaesthetist, LTHTR, Preston, UK. During his career Dr Khan has developed a passion for medical education and acquired higher qualification in this field alongside his parent speciality in anaesthetics. He has held clinical academic posts, first at the University of Oxford and then at the University of Manchester. He was Associate lead for 'Assessments' at the Manchester Medical School. He is currently working in the UAE as a Consultant Anaesthetist. He has extensively presented and published at the national and international levels in his fields of academic interest.

DR KATHRYN GAUNT, MBChB, MRCP, PGD(MedEd), at the time of production of this manuscript was a Medical Education and Simulation Fellow, LTHTR, Preston, UK. Dr Kathryn Gaunt graduated in 2002 from the University of Manchester Medical School, UK. She was a Specialty Registrar in Palliative Medicine and has a keen interest in medical education. She is currently pursuing a higher qualification in this field. She has recently taken time out of her training programme to work as a Medical Education and Simulation Fellow at the Lancashire Teaching Hospitals NHS Trust. Here she lectured, examined for OSCEs and was involved in research and development within the field. She has now returned to her training in Palliative Medicine.

DR SANKARANARAYANAN RAMACHANDRAN, MBBS, PgCert(MedEd), FHEA, is a Medical Education and Simulation Fellow, Lancashire Teaching

Hospitals NHS Trust, Preston, UK. Dr Ramachandran qualified in India and underwent postgraduate training in General Medicine. He had been teaching and training undergraduate and postgraduate medical students throughout his career. He has completed a Postgraduate Certificate and pursuing a Master's degree in Medical Education at the University of Manchester. He has special interest in assessment, problem based learning and research in medical education.

Dr PIYUSH PUSHKAR, MBChB, works in Critical Care at the Aintree University Hospitals Trust, Liverpool, UK. Piyush qualified from the University of Edinburgh and has subsequently worked and trained in anaesthesia and critical care in Lancashire and Merseyside, as well as aeromedical retrieval medicine in Queensland, Australia. He also works as a volunteer doctor for Freedom from Torture in Manchester.

References

- Adamo G. 2003. Simulated and standardized patients in OSCEs: Achievements and challenges 1992–2003. *Med Teach* 25:262–270.
- Angoff WH. 1971. Scales, norms and equivalent score. In: Thorndike RL, editor. *Educational measurement*. 2nd ed. Washington DC: American Council on Education. pp 508–600.
- Barrows SH. 1993. An Overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med* 68:443–451.
- Bloch R, Norman G. 2012. Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Med Teach* 34:960–992.
- Cleland JA, Abe K, Rethans J-J. 2009. The use of simulated patients in medical education: AMEE Guide No 42. *Med Teach* 31:477–486.
- Cohen-Schotanus J, Van Der Vleuten CP. 2010. A standard setting method with the best performing students as point of reference: Practical and affordable. *Med Teach* 32:154–160.
- Cohen R, Reznick R, Taylor B, Provan J, Rothman A. 1990. Reliability and validity of the objective structured clinical examination in assessing surgical residents. *Am J Surg* 160:302–305.
- Collins JP, Harden RM. 1998. AMEE Medical Education Guide No. 13: Real patients, simulated patients and simulators in clinical examinations. *Med Teach* 20(6):508–521.
- Cusimano MD, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R. 1994. A comparative analysis of the costs of administration of an OSCE (objective structured clinical examination). *Acad Med* 69:567–570.
- Downing SM. 2003. Item response theory: Applications of modern test theory in medical education. *Med Educ* 37:739–745.
- Ebel R. 1972. *Essentials of educational measurement*. New Jersey, NJ: Prentice-Hall.
- Epstein RM. 2007. Assessment in Medical Education. *N Eng J Med* 356:387–396.
- Friedman Ben-David M. 2000. Standard setting in student assessment. Association for Medical Education in Europe.
- GMC. 2009. Tomorrow's Doctors [Online]. London: GMC. [Accessed 10 June 2012] Available from http://www.gmc-uk.org/static/documents/content/GMC_TD_09_1.11.11.pdf
- Harden RM, Stevenson M, Downie WW, Wilson GM. 1975. Assessment of Clinical Competence using Objective Structured Examination. *BMJ* 1:447–451.
- Hodges B, McIlroy JH. 2003. Analytic global OSCE ratings are sensitive to level of training. *Med Educ* 37:1012–1016.
- Humphrey-Murto S, Smee S, Touchie C, Wood TJ, Blackmore D. 2005a. A Comparison of Physician Examiners and Trained Assessors in a High-Stakes OSCE Setting. *Acad Med* 80:S59–S62.
- Humphrey-Murto S, Wood TJ, Touchie C. 2005b. Why do physicians volunteer to be OSCE examiners? *Med Teach* 27:172–174.
- Humphris GM, Kaney S. 2001. Examiner fatigue in communication skills objective structured clinical examination. *Med Educ* 35:444–449.
- Kaufman DM, Mann KV, Muijtjens AM, Van Der Vleuten CP. 2000. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. *Acad Med* 75:267–271.
- Khan K, Ramachandran S. 2012. Conceptual Framework for Performance Assessment: Competency, Competence and Performance in the Context of Assessments in Healthcare – Deciphering the Terminology. *Med Teach* 34:920–928.

- Kowlowitz V, Hoole AJ, Sloane PD. 1991. Implementing the objective structured clinical examination in a traditional medical school. *Acad Med* 66:345–347.
- Kramer A, Muijtjens A, Jansen K, Dusman H, Tan L, Van Der Vleuten C. 2003. Comparison of a rational and an empirical standard setting procedure for an OSCE. *Objective structured clinical examinations. Med Educ* 37:132–139.
- Mann KV, Macdonald AC, Nornici JJ. 1990. Reliability of objective structured clinical examinations: Four years of experience in a surgical clerkship. *Teach Learn Med* 2:219–224.
- Mccoy JA, Merrick HW, editors. 2001. *The Objective Structured Clinical Examination*. Association for Surgical Education. Springfield, IL.
- Morgan PJ, Cleave-Hogg D, Guest CB. 2001. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Acad Med* 76:1053–1055.
- Newble D. 2004. Techniques for measuring clinical competence: Objective structured clinical examinations. *Med Educ* 38:199–203.
- Newble DI, Hoare J, Sheldrake PF. 1980. The selection and training of examiners for clinical examinations. *Med Educ* 14:345–349.
- Pell G, Fuller R, Homer M, Roberts T. 2010. How to measure the quality of the OSCE: A review of metrics-AMEE guide no. 49. *Med Teach* 32:802–811.
- PMETB. 2007. Developing and maintaining an assessment system-a PMETB guide to good practice [Online]. London: PMETB. [Accessed 10 June 2012] Available from http://www.gmc-uk.org/assessment_good_practice_v0207.pdf_31385949.pdf
- Reznick RK, Regehr G, Yee G, Rothman A, Blackmore D, Dauphinee D. 1998. Process-rating forms versus task-specific checklists in an OSCE for medical licensure. *Medical Council of Canada. Acad Med* 73:S97–S99.
- Roberts C, Newble D, Jolly B, Reed M, Hampton K. 2006. Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Med Teach* 28:535–543.
- Shavelson R, Webb N. 2009. Generalizability theory and its contributions to the discussion of generalizability of research findings. In: Ericikan K, Roth W-M, editors. *Generalizing from educational research: Beyond qualitative and quantitative polarization*. New York, London: Routledge. pp 13–32.
- Shumway JM, Harden RM. 2003. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach* 25:569–584.
- Stevens DD, Levi A. 2005. Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning. Sterling, VA: Stylus Pub.
- Stillman PL. 1993. Technical Issues: Logistics. *Acad Med* 68:464–468.
- Tan CP, Azila NM. 2007. Improving OSCE examiner skills in a Malaysian setting. *Med Educ* 41:517.
- Tavakol M, Dennick R. 2011. Post Examination Analysis of Objective Tests: AMEE Guide 54. *Med Teach* 33:245–248.
- Tavakol M, Dennick R. 2012. Post-examination interpretation of objective test data: Monitoring and improving the quality of high-stakes examinations: AMEE Guide No. 66. *Med Teach* 34:e161– e175.
- Turner JL, Dankoski ME. 2008. Objective Structured Clinical Exams: A Critical Review. *Fam Med* 40:574–578.
- Van Der Vleuten CPM, Van Luyk SJ, Van Ballegooijen AMJ, Swansons DB. 1989. Training and experience of examiners. *Med Educ* 23:290–296.
- Vargas AL, Boulet JR, Errichetti A, Zanten MV, López MJ, Reta AM. 2007. Developing performance-based medical school assessment programs in resource-limited environments. *Med Teach* 29:192–198.
- Whelan GP. 1999. Educational commission for Foreign Medical Graduates – clinical skills assessment prototype. *Med Teach* 21:156–160.
- Wilkinson TJ, Frampton CM, Thompson-Pawcett M, Egan T. 2003. Objectivity in Objective Structured Clinical Examinations: Checklists Are No Substitute for Examiner Commitment. *Acad Med* 78:219–223.
- Williams RG. 2004. Have Standardized Patient Examinations Stood the Test of Time and Experience? *Teach Learn Med* 16:215–222.

Appendix 1

Observed OSCE station Filled Template

Question Info

Author

AQ

Subject/Topic

Asthma

Level of the Candidate (choose at least 1)

☒ Year 3☐ Year 4☒ Year 5

Competencies (Essential Field)

Please choose 1 to 3

☒ Clinical Examination Skills☐ Communication & Consultation Skills☒ History Taking☐ Hand Over Skills☐ Procedural Skills☐ Mental Health Assessment☐ Professionalism☒ Application of Knowledge☐ Prescription☐ Data Interpretation☐ Diagnostic Skills☐ Management Planning☐ Ethics

Station Duration (please do not modify this field)

8 min

Information for the site organisers

SP age & sex

Mark/Mary Freeman, age 32

Resources & Equipment needed

1. Paper and pen, in case the candidate wishes to make notes
2. Alcogel™
3. Couch for the patient
4. Stethoscope
5. Water bottle and glasses

Setting up the station

1. Examiner's chair should be positioned so that he/she can observe faces of both candidate and simulated patient.
2. No desk or table is necessary. If one is present, it should NOT form a barrier between the candidate and the patient.

Instructions for Candidates (outside the station)

You are a medical student, on your placement at the GP Practice

Mark/Mary Freeman is a 32-year-old patient who is attending for their annual asthma review. This is the first time that Mr/Mrs Freeman has attended this year.

- (1) You are expected to take a brief, focused asthma history from this patient to assess his/her asthma control.
- (2) Perform a focused respiratory system examination

Please do not take a detailed history

Please do not perform a general physical examination

Information for the examiner**Brief Background to the scenario**

Mark/Mary Freeman is a 32-year-old patient who is attending for their annual asthma review

Examiner's Role

Your role is to observe the history taking process and to assess the candidate's examination of the respiratory system

What are the objectives of the station or what is expected of the candidate?

- The candidate is expected to take a brief, focused asthma history from this patient to assess his/her asthma control and make to suggestions as appropriate
- The candidate is also expected to examine this patient's chest

The candidate has NOT been asked to take a detailed history or perform a general physical examination

What information they might be able to provide the candidate?

- If the candidate attempts to perform a general physical examination, please ask them to move on to examining the chest
- If candidates attempt to perform tactile vocal fremitus and whispering pectoriloquy please ask them to omit these and move on

What information they should not provide the candidate?

- Please do not repeat instructions to the candidate
- Please do not interrupt or prompt candidate or ask any questions

Clinical information relevant to the station

Key points in Asthma history taking;

1. Onset
2. Duration
3. Family History
4. Occupation
5. Smoking history
6. Housing, soft toys
7. Aggravating and relieving factors, e.g. exercise, weather, allergies, etc.
8. Cough
9. Wheeze
10. Perennial versus Seasonal
11. Diurnal variation/night time symptoms
12. Affect on lifestyle
13. Hospital admissions
14. Treatment including steroids

Key Points in Examination of the Respiratory System;

(A good student would complete the examination at the front before moving to the back)

1. Proper exposure
2. Inspection for symmetry and shape
3. Palpation of trachea, chest expansion (vocal fremitus not necessary in this case)
4. Percussion, front and back including the clavicles
5. Auscultation at front and back, (whispering pectoriloquy not necessary in this case)

British Thoracic Society (BTS) guidelines

BTS/Sign Guidance on pets in asthma: There are no controlled trials on the benefits of removing pets from the home. If you haven't got a cat, and you've got asthma, you probably shouldn't get one.

BTS/Sign Guidance on smoking and asthma: Direct or passive exposure to cigarette smoke adversely affects quality of life, lung function, need for rescue medications and long-term control with inhaled steroids.

BTS/Sign Guidance on control of asthma: The aim of asthma management is control of the disease. Complete control is defined as:

- no daytime symptoms
- no night time awakening due to asthma
- no need for rescue medication
- no exacerbations
- no limitations on activity including exercise
- normal lung function (in practical terms FEV1 and/or PEF >80% predicted or best)
- minimal side effects from medication

Simulated Patient Information**Who they are?**

Mark/Mary Freeman, age 32

Their social/economic background if applicable?

- You are currently working in a telesales call centre (trying to sell fitted kitchens)
- You live in your own small terraced house, with your partner and one daughter
- You bought your daughter a cat 4 months ago (for her birthday)
- Your partner smokes 20 a day and you smoke occasionally – when you go out socialising. You know you shouldn't, but you don't think an occasional one harms your chest. You've been doing this since you were 16
- You have 2–3 drinks (pints of beer or large glasses of wine) when out socialising – which is once a week or less often
- You go to the gym very occasionally
- Your mood is normal (not anxious or depressed)

History

- You have had Asthma most of your life – you think it was diagnosed when you were about 5 or 6
- It seemed to be much worse as a child – you had been under a hospital clinic, but you have never had to stay in hospital
- As a teenager, it seemed to go much better – you rarely seemed to need treatment
- In your adult years, it hasn't troubled you too much, though you have always liked to have a blue inhaler (salbutamol) handy, just in case
- If asked, you have never taken a course of steroid tablets by mouth and you do not own your peak flow meter
- One inhaler usually lasted you several months, as you only needed it every few days (but see below for current situation). When you use it, you take 2 puffs at a time (as instructed), which relieves the coughing and wheezing noises within 5 min or so)
- You have an elder brother with asthma, a your daughter has mild eczema and your father has had hay fever all his life
- Your asthma is not seasonal
- It does not vary by day or night

Details of their concerns/perceptions

- You also wonder if the cat might be to blame. On the other hand, you had a dog when you were young, which didn't affect you, so you don't think you are allergic to animal fur
- You are not really worried about your asthma – you don't think it's anything serious, but you're happy to come for a check-up

What they should say (their agenda) & what they should not say

You think the reason why your asthma is worse is because of working longer hours. There is a fair amount of stress at work, as the company's sales are not doing well. You don't feel that you are getting over-stressed; it's just the general atmosphere at work

What they should ask (questions)

Please do not ask any questions

Specific Standardisation issues (specific answers to specific questions, please stay with the script)

If the candidate asks for your thoughts about why your asthma has worsened very early on (before making any effort to establish rapport), say you believe it's the stress at work. But if the candidate asks the same question when you have established some rapport, mention your belief about it being the stress at work, but also say: *It's been ever since my daughter's birthday, I suppose.* If they probe why you think that might be, mention the cat.

Appendix 2

OSCE QA Questionnaire

Question Title/Number

Feasibility Question

How easy will it be to find an SP for this question?	Very easy	Relatively easy	Difficult
--	-----------	-----------------	-----------

Validity Questions

Are the tasks in this station achievable in 8 Minutes?	Yes	No
Is this topic taught in curriculum?	Yes	No
Where is it taught in curriculum?		

The attributes being tested are expected of a FY1 doctor.	Yes	No
Would this question discriminate between good and poor students?	Yes	No
Would this station be able to recreate an atmosphere close to real patient encounter?	Yes	No
If a candidate passes this station would that be able to extrapolated to competence in workplace?	Yes	No
Does this question test what it is supposed to test – as outlined in primary and secondary competences?	Yes	No

Supplemental Questions

Would this question benefit from piloting if unsure about any of the above?	Yes	No
Can this question be used in other years as well?	Yes	No
If yes please indicate which year		