Objective Structured Clinical Examination

OSCE MANUAL 2014

الهيئة السعودية
للتخصصات الصحية

Saudi Commission for Health Specialties

Prepared by:
Professor James Ware
Professor Abdelmoniem El Mardi
Dr Hamza Abdulghani
Dr Imran Siddiqui

Formatted by:
Vittoriana Crisera

King Fahd National Library Cataloging-in-Publication Data

Saudi Commission for Health Specialties, Riyadh - 2014
Objective Structured Clinical Examination

33p; 14.8cm × 21cm

ISBN: 978-9960-9832-8-8

1-Clinical medicine – Diagnosis I-Title

616.075 dc 1435/9119

L.D. no. 1435/9119

ISBN: 978-9960-9832-8-8

TABLE OF CONTENTS

Preface

The Saudi Commission for Health Specialties bears the responsibility for both postgraduate education and assessment standards of specialist training. Following the introduction of postgraduate training in the Kingdom over twenty years ago, the Commission is now undertaking a major overhaul of the programmes offered, the training delivered, their accreditation and finally and very importantly the testing of candidates.

There are many challenges ensuring the right standards of care given by our newly certified specialists and future consultants. In this respect the Commission would like to acknowledge the enormously important roles that all our professional colleagues contribute with.

It is no secret that there is no one absolute process that can and should be applied to the process of assessment. In this manual often several different solutions are offered for the same purpose and in these cases the individual exam committees will be free to choose their own solutions. The Commission's role is to provide the overall policies and regulations set out in the General Examination Rules and Regulations handbook which is updated and published periodically. We would commend all those who read this manual to read those regulations.

The commission has now built three state of the art OSCE centres in Riyadh, Dammam and Jeddah and it is hoped this will encourage all those participating in the important task of examining clinical competence.

Finally I would like to add my sincere words of thanks to those who have contributed to writing this clear and informative manual.

Professor Abdulaziz Al Saigh
Secretary General
The Saudi Commission for Health Specialties
Riyadh

Introduction

This manual describes how to plan, construct and deliver an Objective Structured Clinical Examination (OSCE). An OSCE is an examining process made up of a series of stations of equal length set in a circuit. A postgraduate OSCE usually has 10-15 stations each of eight to fifteen minutes. The OSCE is considered to measure clinical competence and was first created by a group of senior examiners in Glasgow, UK, to replace the undergraduate finals' long case, which is inherently unreliable. Since that time the OSCE has been used in postgraduate and national licensing examinations in the USA, UK, Canada and Australia among other countries. The Saudi Commission for Health Specialties has chosen the OSCE to test clinical competence for the postgraduate medical Saudi Board Examinations. The Commission has also built three OSCE centres in Riyadh, Jeddah and Dammam. It is important that assessment tools are used for their correct purpose, and for example, a structured VIVA which some confuse with an OSCE, is used to assess reasoning, critical thinking, judgment and application of knowledge. The OSCE is a practical test of medical, or surgical, obstetric and so on, practice. In other words what a specialist will actually do. The Structured VIVA is a theoretical concept and is much more a test of a candidate's intellectual ability.

The OSCE is very resource intensive and should not be undertaken by those without experience. This manual is for examiners with some experience of running OSCEs. A second and equally important reason for the manual is to ensure that OSCEs used to certify specialist competence meet the standards required for a valid, credible and defensible test.

In the chapters that follow there are details for the hands on construction of a testing process that will meet the Commission's standards for testing postgraduate clinical competence. Examples of what will meet these standards include the use of a test blueprint, provision of written materials to guide examinees, examiners, patients and simulated patients, equipment lists and grading forms. The planning of an OSCE must begin many months before the delivery date. This will allow for the choosing of examiners, contacting and training of simulated patients, collection of exam materials such as X-rays and common medical equipment required to make stations authentic tests of what is supposed to be measured, for example male catheterization. Training or orientation cannot and should not take place on the same day as the OSCE itself, instead preferably about one week before, so that those being trained do not forget what has been said.

There has been considerable evidence reported from research studies which can advise how stations should be run. For example, initially it was thought that two

examiners at a station would be more reliable than just one. The evidence clearly shows that it is better to use all second examiners to have more stations and thereby increase the reliability of the whole examination. The way examiners grade candidates is another important issue and whatever details are used it must ensure every candidate is assessed in the same way. In fact this is another characteristic of the OSCE process that all candidates do exactly the same examination, something that could hardly happen with the long case.

Any assessment shall be fit for purpose, in other words an assessment of clinical competence at medical school graduation level shall determine that successful candidates meet the minimum standards for safe practice. A test of clinical competence at Saudi Board Part II level is much more about standards of excellence commensurate with the practice of a specialist. It serves no purpose if the OSCE designed for a Saudi Board Part II surgical exam only tests at the level of a graduating medical student, where the core undergraduate curriculum consists mainly of the study of common conditions. A resident is expected to have in depth knowledge of their specialty including unusual conditions. Thus when constructing an OSCE for a Saudi Board Part II exam, counseling a simulated patient over common conditions, such as before an appendicectomy, would probably not be a suitable choice for a station, instead counseling a patient before open heart or parathyroid surgery would make more appropriate tests of clinical competence. However, the choice of stations stems from the published competencies presented as outcomes of the curriculum/ training participated in.

Finally, it will be appreciated by reading this manual that to create and deliver a successful OSCE requires many administrative duties and so those responsible must have these resources to ensure success. All members of the OSCE team will be pleased to assist any group who have been tasked with the duty of constructing an OSCE according to the Saudi Commission standards.

The Saudi Commission OSCE Team

Professor James Ware	Saudi Commission for Health Specialties	j.ware@scfhs.org
Professor Abdelmoniem El Mardi	College of Medicine – Al-Imam University	elmardi@gmail.com
Dr Hamza Abdulghani	College of Medicine – King Saud University	hamzaabg@gmail.com
Dr Imran Siddiqui	Saudi Commission for Health Specialties	i.siddiqui@scfhs.org

Chapter One: Relevant principles of assessment

1.1 Clinical competence

Competence

A Proper definition of clinical competence and its components is important to serve as a criterion for validating medical educational programs and to assure a minimum level of competency at the end of medical school and beyond during residency. Webster's dictionary defines being competent as the quality of having sufficient knowledge, judgment, skill, or experience for some purpose. A distinction between competence and performance is often made in the literature. Competence is the possession of the requisite or adequate ability, having acquired the knowledge and skills necessary to perform those tasks that reflect the scope of professional practice. It may be different from performance, which describes what someone is actually doing in a real life situation. Senior (1976), for example, defined competence as what a physician is capable of doing and performance as what a physician actually does.

Competencies are a set of professional abilities that includes elements of knowledge, skill, attitudes and experience.

Clinical competence

A trainee's ability to do what is expected at a satisfactory level of delivery, at a certain point in time, e.g. at the completion of the program. It is the acquisition of a body of relevant knowledge and of a range of relevant and related skills which includes personal, interpersonal, clinical and technical components. In the case of clinical education, which is still primarily based on an apprenticeship model, teachers define what the trainee is expected to do and then test their ability to do it. However, because of the complex reality of what doctors actually do on a day-to-day basis, 'clinical competence' gives a rather limited view of their work, professional experience and expertise. Most clinical actions are concerned with problems for which there is no clear answer or no single solution and where no two patients are the same, even if they have the same condition. An experienced doctor searches his or her mind and sifts through a wide range of options and in some cases the solution will be something he or she has never come up with before. Therefore, competence itself is best seen as a prerequisite for performance in the real clinical setting where it would be expected that a doctor operated at a higher level in many areas and demonstrated mastery in some.

Competence encompasses a variety of competencies which are divided differently into a number of categories, themes, outcomes or attributes by different professional bodies.

Competence as an assessment term

George E. Miller (1990) distinguished several hierarchical layers of competence to function as a framework within which assessment might occur. The different layers in Miller's model represent a developmental sequence of stages.

Miller conceived competence as a pyramid. The base of the pyramid consists of factual knowledge. One level up, Miller describes the ability to use knowledge in a particular context as "knows how." This comes close to clinical reasoning and problem solving. At a higher level, "shows how" reflects the person's ability to act appropriately in a practical situation and describes hands-on behaviour in a simulated or practice situation. The "does" level refers to actual performance in habitual practice. The higher the skills being tested in the pyramid, the more clinically authentic the assessment needs to be.

In the field of assessment in medical education, the term competence means the ability to apply clinical skills in a standardized setting, the examination setting. On the other hand the term performance means the ability to apply clinical skills in a real life clinical situation. Tests of performance carried out in a clinical setting are called workplace-based assessment. For the purpose of assessment of clinical competence (assessment of skills mainly), the following simple classifications were found to be most meaningful:

(1) Communication skills

These skills lead to proficiency in communication - an essential skill for clinical practitioners because of the large and varied number of people doctors must communicate with every day and the range of circumstances, some of which might be very distressing. The idea that doctors automatically learn communication through experience or that doctors are inherently either good or bad communicators is long abandoned. It is now widely acknowledged that both students and postgraduate doctors can be educated in communication skills and their proficiency can develop to extremely high levels of expertise.

1.a. Communication skills for history taking

These are the skills applied during the interviewing of patients for the purpose of information gathering. The communication skills range from the ability to follow the

agreed consultation framework to the actual use of certain communication skills to initiate, build and maintain rapport with the patient and thereby facilitate the process of information gathering. This competency is usually acquired at two levels. The first level is the ability to apply the techniques in the consultation framework and the second level is the ability to take a focused history and analyse the patient responses.

1.b. Communicating with patients in other situations

These are the protocols for skills used during situations other than history taking, and include:

- When the consulting practitioner admits mistakes or errors
- Discussion for the purpose of shared decision-making
- Discussions with patients for the purpose of advising or counseling
- Disclosure of information voluntarily or in response to patients' questions, when conveying bad news or getting informed consent

1.c. Communicating with others

These are the skills used when practitioners communicate with patients' relatives, colleagues or the public including authorities and the media

1.d. Communicating in writing and electronically

These are the skills needed by practitioners to enable them to communicate in writing and electronically with colleagues and others

(2) Physical examination skills

2.a. Basic Physical examination techniques:

These are the skills used by practitioners when they examine patients with various clinical conditions. At an early stage of medical education, learners are expected to acquire and master the techniques using basic clinical skills.

2.b. Advanced physical examination techniques

Later in their training, learners and trainees are expected to be able to use physical examination skills to examine patients and be able to identify, elicit and interpret the major signs of common clinical conditions.

Trainees are expected to know and apply the special tests and techniques like the straight leg raising test, the tendon reflexes, Valsalva manoeuvre and so on, and to elicit signs like Babinski, Rovsing and clonus.

2.c. Virtual sign identification

As it is now customary to use multimedia learning resources and simulations in the teaching of clinical conditions, it is expected that residents are able to identify clinical signs when they encounter them in a virtual reality setting such as with videos, audio, computer and physical simulations. Investigations were conducted to study the predictive validity of virtual identification on the ability of residents to identify these signs when they encounter them in real patients. Results indicated that there is high correlation between the two.

(3) Procedural skills

3.a. Therapeutic procedures

Residents are expected to learn and acquire a certain level of mastery for some major therapeutic procedures like First Aid and Basic Life Support procedures, intubation, application of catheters, injections, suturing, reductions of dislocations and so on. They are also expected to be able to operate and monitor certain therapeutic and monitoring devices and to apply a specified range of therapeutic appliances.

3.b. Diagnostic procedures

Residents are expected to learn and acquire a certain level of mastery for some major diagnostic procedures such as venepuncture, lumbar puncture, fluid aspiration, electrophysiological testing (ECG, EMG), and a range of laboratory and bedside diagnostic tests. They are expected to learn about and be familiar with, but probably not apply some important diagnostic procedures like imaging, histopathological examinations and molecular biological investigations.

3.c. Identification and interpretation of abnormal test findings

Postgraduate trainees are expected to perform all important diagnostic procedures, as well examine results and identify and interpret those that are abnormal.

(4) Cognitive skills

4.a. Data interpretation (knowledge application)

Residents are expected to interpret the data collected during patient interviews to guide them through the physical examination process and interpret the history and examination data to decide on a plan of management. The examinee would then be expected to interpret the pattern of all the investigations and build a picture of the condition the patient is suffering from and reach a definitive diagnosis.

4.b. Decision making

During the ordinary clinical work-up, practitioners are expected to be continually making diagnostic, therapeutic and ethical judgments. Therefore decision making abilities are of paramount importance for clinicians.

4.c. Problem solving

Through data interpretation and decision making skills, medical students acquire the ability for problem solving.

1.2 Assessment of clinical competence

According to the general classification of learning's three major domains, namely knowledge (cognitive), skills (psychomotor), and attitude (affect), medical schools use three different types of tests:

- Achievements tests that mainly focus on assessments of knowledge.
- Tests of performance (competence) that mainly concentrate on the assessment of skills.
- Tests of professionalism, which mainly concentrate on the assessment of communication, ethics and behaviour.

There is, however, an unavoidable overlap between the three domains of learning and therefore care should be taken that during testing, emphasis should not be allowed to shift significantly from the intended domain to the overlapping domains.

Elstein and his colleagues concluded from their studies in 1978, that clinical competence is highly dependent on the particular content of the situation and they called this phenomenon content specificity. According to other studies the ability to solve clinical problems did not seem to be a general, content-independent characteristic of doctors. The reasoning that clinical competence is more than solving clinical problems was replaced by a perspective that knowledge is an essential factor in all competencies (Van der Vleuten et al., 2000).

Tests of performance

Various types of tests are used to assess the skill components of clinical competence. Users of the various types have their justifications for using any particular approach. It is generally agreed that for an institution to have a fair and comprehensive assessment of clinical competence, a battery of diverse testing approaches should be used.

1.3 Characteristics of a good assessment instrument

To judge the utility of any particular assessment instrument, it is necessary to apply the following criteria:

a. Objectivity

An objective instrument is one that assessors can easily agree on what is required from an examinee as a response or performance when they read the item and what constitutes an acceptable (correct) response or performance for which the assessors assign a grade. It must be remembered, however, that objectivity in not simply the reduction or abolition of examiner subjectivity. When the chance given to examiners to manipulate the questions, answers, responses, performance and judgment are minimized, the chances of objectivity are greatest.

b. Validity

This is defined as the extent to which a test measures what it purports to measure. There is overwhelming evidence that the reliability of measurements of clinical competence is reduced by the fact that competence is content specific.

Achieving competence in one area (for example, in one clinical case) is not a good predictor of competence in another, even if the areas are closely related. There are two main types of validity. The extent to which a test is suitable for the task (fit-for-purpose) is called its face validity. The extent to which the test(s) cover(s) the subject, discipline or domain is called its content validity. When an examination is carefully designed through good selection and weighting of the topics to be assessed it is described as having content validity.

c. Reliability

This is also called repeatability, reproducibility or consistency. It is the extent to which a test or the instrument if repeated for the same or similar group of examinees will produce the same or similar results. A reliable instrument should be both accurate and precise. So reliability refers to the precision of measurement or the reproducibility of the scores obtained with the examination. Reliability means reproducibility of scores across raters, questions, cases, occasions and should be capable of differentiating consistently between high and low ability students. Because competencies are highly domain-specific, broad sampling is required to obtain adequate reliability across content for example, range of cases/situations, across other potential factors that cause error variance for example with testing time, examiners, patients, settings, facilities. Certain factors are known to improve reliability like objectivity and validity.

e. Educational Impact

Examinations drive learning; this hypothesis describes one of the strongest relationships in education. Trainees wish academic success, and academic success is defined by examinations, therefore students will do everything to maximize their chances of success. The way the exam is conducted also affects the way students learn. An instrument that drives students to learn more in depth is said to have high educational impact.

f. The utility index

The 'utility index', described by Cees van der Vleuten in 1996, serves as an excellent framework for assessment design and evaluation .

Utility = Educational impact x validity x reliability x cost x acceptability

It acknowledges that optimising any assessment tool or programme is about balancing the six components of the utility index. Choice of assessment instruments and aspirations for high validity and reliability are limited by the constraints of feasibility, such as the resources to deliver the tests and the acceptability to the trainees. For example, a high stakes examination on which progression into higher specialist training is dependent will need high reliability and validity, and may focus on this at the expense of educational impact. In contrast, an assessment which focuses largely on providing a trainee with feedback to inform their own personal development planning would focus on educational impact, and less emphasis on reliability.

1.4 The objective structured clinical examination (OSCE)

The OSCE was first described by Harden in the 1975 in an attempt to overcome the problems associated with traditional methods for examining clinical competence, such as with unobserved long and short cases and unstructured VIVAs. What was described was an organizational framework incorporating a variety of test methods. It incorporates the possibility of using a number of different assessment methods within the single examination. As a new evaluation tool that allowed trainees to be observed performing in many different clinical situations, the OSCE was a major advance over traditional methods in which only one clinical encounter was examined but most often not observed.

The OSCE also incorporated the use of standardized patients first described by Barrows and Abrahamson in 1964. The use of standardized patients allowed the nature of problems and the level of difficulty to be standardized for all students (van

der Vleuten and Swanson). This combination of multiple observations and standardization of content and difficulty has made the OSCE a very popular evaluation tool.

Further, extensive research demonstrated that OSCEs could have excellent psychometric properties. As a result, the use of OSCEs is now extensively used in medical schools throughout the world. OSCEs have become indispensable for the assessment of medical students, clinical clerks, interns, and residents as well as candidates for licensure and certification. OSCEs are also widely used for the assessment of the competence of other health professionals, including chiropractors, nurses, nurse practitioners, pharmacists, dentists, and physiotherapists. As described by Harden and Gleeson in 1979, the OSCE is a series of stations usually all of equal length, 5-10 minutes, at which various competencies are tested using manned and unmanned stations.

Importantly, the measurement is structured, using specially prepared mark sheets with specific instructions and training provided for the examiners. A time allowance may be given for candidates to move between stations and often REST stations are interspersed in the circuit to allow candidates to prepare themselves for the coming stations and complete any incomplete questionnaire. The candidate's answers from unmanned stations may be carried in an OSCE book, while examiners record their marks on separate sheets which are not given to the candidate.

If the OSCE is to attain reliability there must be sufficient stations to neutralise the effect of case specificity. It is a timed examination in which examinees move from station to station; each station requiring performance in a simulated setting that usually involves interaction with a standardized patient. The examinee is typically required to demonstrate some combination of history taking, physical examination, counselling, or other aspect of patient management. At each station, candidates' performances are rated on checklists and global rating scales. OSCEs have been extensively studied and established as performance-based assessment instruments with good validity and reliability. Indeed, because OSCEs have been shown to have a much higher reliability and validity than traditional and less structured oral examinations, they have grown in acceptance as a means of episodic performance-based assessment. The traditional assessment methods for clinical competence were seen as being unreliable and often lacking validity.

Although having more stations is better than less, there is no absolute number even for a high stakes examination. However, when adding the rest stations this may be more preferable than using multiple circuits. This is not an uncommon solution to

accommodate candidate numbers at large schools with many candidates to cope with rather than run the same examination on separate days for different sets of candidates.

The use of paired examiners is no longer recommended and instead a larger number of stations using all examiners as single observers at the additional stations is preferred and shown to give better reliability. It is desirable to know how long a borderline candidate might take and determine station length on this basis.

Best of all, of course, is to pre-test stations or know from previous editions of the examination how long chosen competencies take to test effectively. This can be determined using formative OSCEs with a limited number of stations.

The OSCE is very flexible and can test a wide range of tasks and skills. The choice should and will always be within the boundaries set by stated outcomes and more specific learning objectives. Consideration also needs to be given to whether or not the candidates are familiar with the format of certain “unusual” stations otherwise the danger of exam by ambush becomes a problem. To avoid this, again formative OSCEs even if not full length, are to be highly recommended.

Another important issue to address is to satisfy the need for a wide variety of skills and competencies by using a test blueprint or matrix to design the OSCE itself. In this respect, the other exam components should not be forgotten, such as paper tests of knowledge, and other higher tests of cognition such as MEQs. If the whole examination presents an unbalanced and overlapping distribution of competencies tested the value of the OSCE is diminished.

The choice of tasks is part of the organizational procedure for setting up a successful OSCE and above all shall reflect the core competencies expected of the students. The organizing team needs also to consider whether an OSCE can reasonably give information about all attributes expected from the candidates at the level of the test or indeed be possible to test at all.

Chapter Two: Item Construction

Once a station list is compiled, the stations can either be pulled from an item bank or prepared by members of the OSCE committee. Members can distribute the stations among themselves or ask other staff to participate in the preparation process.

All station designers need to prepare a number of documents related to their station.

The process is best started in a meeting where the participants agree on what needs to be prepared and then decide a time for the review meeting of all the stations. The process is the same for active stations - where candidates are expected to interact with a patient or simulator, and static stations - where candidates are expected to carry out an unobserved task and then answer questions based on their findings. A station designer or constructor needs to prepare the following, taking into consideration that the products are intended for upload into an item bank and not the exam:

2.1 Station's profile

This profile is usually used for the identification of the characteristics of the station in the item bank. It includes the condition (such as a case of hyperthyroidism) presented in the station, usually used as a title for the station. An estimate of the time needed for the task, and the domain and competence tested (objective/construct of the station). The profile should also include a proposal of the suitable examiner (for example a surgeon). A paramedic would be quite sufficient for a station rating a medical student or resident demonstrating CPR. The documentation should also specify the item which is most suitable for the demonstration of the skill, whether it is going to be a patient (non-standardised or standardised), a model /mannequin, a simulator (physical or digital) or an interactive animation and so on. If the item is not a patient, the alternative item should also be clearly specified, for example a video of a patient with respiratory distress and audible wheezes. Station materials must be prepared in good time.

2.2 Station's opening statement (scenario, preamble)

This is important as it is expected to prepare the mindset of the examinee and provide a semi-real-life situation. A scenario should, as much as possible, use the

patient's language rather than technical medical terminology. It should provide enough information to guide the examinee to the required task, but not too little. A standardized format for the "stem" (task) is helpful - for example, providing the patient's name, age, presenting complaint, and the setting (such as clinic, emergency department, or ward) - and must be uniform for all stations. Bearing in mind the time constraint for any station, care must be taken not to overload the scenario with any unnecessary detail.

2.3 Instructions

Clear instructions should be written for examiners, examinees and patients (non-standardised or simulated/standardized).

a. Instructions for candidates

Candidate instructions may state the specific task or skill to be demonstrated. They must be clear and precise, whereby any person reading the instructions will have a full understanding of the task. It is preferable that only one task is required per station, such as demonstration of a thyroid examination at undergraduate level and demonstration of the examination of a patient with thyroid disease for a resident. At an undergraduate level the instruction to auscultate the heart of this patient is better than examine the heart of this patient, which is in turn better than examine the cardiovascular system of this patient.

At integrated stations, more than one competence can be tested, in which case, clear instructions for both tasks must be provided. Instructions should also include any cognitive components like interaction with the examiner, the answering of questions, commenting on the actions performed and reasoning. Finally, it is always useful for the candidate to know the division of marks between tasks and questions that might be asked.

The candidate instructions are usually presented on one side of laminated A_4; this ensures that the instructions are not damaged by stressed candidates. The instructions shall be clear and unambiguous using font sizes that can be easily read and re-read at a glance.

b. Instructions for the examiners

Explicit instructions should always be provided for the examiners of a station. As an introduction the station profile shall be presented, with a clear statement of the construct being measured. The instructions given the candidate shall also be given so that the examiner may consider any clarifications (to be avoided and sorted out at

the briefing meeting). Of special importance is whether prompting and interaction is allowed and should include when to interact, how it should be done and to what extent. If any modification in the scoring will result from triggering/prompting, that should also be specified.

Full instructions shall be given for marking, for example the one given below for a postgraduate OSCE:

- 2.0/2.0 **Satisfactory** demonstration at required level (specialist/consultant)
- 1.5/2.0 **Borderline** quality
- 1.0/2.0 **Unsatisfactory** due to a significant error or omission
- 0.5/2.0 **Weak** quality for level required
- 0.0/2.0 Multiple **problems** due to significant error(s) or omission(s)

Each level shall be explained in detail with descriptors to ensure that every candidate is assessed in the same way, and this is particularly important when more than one examiner examines the same station (maybe at two or three different centres). In the case of Saudi Board Part II level the required level is that of a specialist/ consultant.

Different tasks may be weighted differently to reflect their importance or difficulty, and so below is a task weighted half that above:

- 1.0/1.0 **Satisfactory** demonstration for required level (specialist/consultant)
- 0.5/1.0 **Weak** quality for level required
- 0.0/1.0 **Unsatisfactory** due to significant error(s) or omission(s)

See below for more details.

c. Instructions and training material for patients

The instructions shall include the most appropriate clothing to be worn and what hospital clothing may be provided to change into. There shall be a short and informative description of what will happen at the station, some stations may

require unusual responses by the SP, such as anger or silence. If a history is to be given the details shall be provided, and preferably surplus to requirements (a complete history) to avoid leading the candidates by the absence of important information that a patient would normally give.
Instructions for non-standardised patients regarding their behavior and interaction with examinees should be clear. Instructions for simulated patients should be detailed and clear and they need to know enough about the condition they will be presenting. A summary of the clinical condition should be provided for all simulated patients. However, paper based instructions are insufficient if a standardised patient is to play the role of a patient with a medical condition - training is required and time must be set aside for this.

If an SP is to be examined, details of the examination shall be given so that the SP will not be taken by surprise. The examiner will most likely carry out the same examination before the start of the OSCE starts so as to determine if there will be any abnormal findings.

NSPs may be used for physical examinations or history taking, but consideration must be given to the number of times they will be put through the same process. Clumsy and painful examination techniques shall result in warnings and large mark deductions (this shall be clearly stated in the examiners' instructions).

If the SP or NSP are going to give grades for their perceptions or experiences of the inter-action with the candidates the instructions must be explicit and also form an important part of the briefings given one week before the examination. These details shall be provided at the end of the instruction sheet, while the grading form is a separate material.

Not infrequently the SP's instructions will run into 3-5 sheets of A_4. Dates, venue and times for briefings shall always be provided, but should not be included for station bank storage.

2.4 Scoring and marking form

The station designer/constructor should follow the rules and standards agreed upon by the examination committee. The format of the form(s) to be used should be specified by the committee as there are various formats. The most widely used are:

a. The checklist

This is a list of the expected items if the task is completely performed. The Committee must decide the degree of detail with which items will be described. Some assessors believe that the statements should be general and flexibility left for the experienced examiner, others believe that statements should be so specific and objective that subjectivity is minimized. The proposed mark for each statement should be included in the checklist.

Statements regarding generic skills and behaviors which are expected to be in most or all stations should not be given a high mark as this will affect the overall mark of the exam. Scoring in a checklist is usually of the yes/no, done/not done type. This method of rating is particularly useful for inexperienced examiners and is not recommended for postgraduate examinations.

b. The rating scale

Like the checklist, a rating scale has a list of statements detailing the task to be performed. The difference is in the scoring style as examiners are interested in differentiating performance based on quality and level of mastery. The candidate will be rated across a scale such as, satisfactory, borderline, weak, not done or problems (see above). A mark is assigned for each level in the scale, like 2, 1.5, 1, 0.5 and 0 for the example given earlier. It is also helpful to provide a descriptor for each level.

Global rating

This concept can be used for different purposes. Some examiners would like to assess the overall performance in the station to differentiate the competent from the insufficiently competent. In this case a mark should be allocated for this global rating. This is usually used with the checklist and is not necessary in the rating scale. Global rating could also be used for the purpose of standard setting in the borderline group or contrasting group methods, in which case tick boxes are provided at the bottom of the scoring form to allow the examiner to pass a global judgment on the examinee indicating whether he considers the candidate a "clear pass", "borderline", or "clear fail". This can be used with both the checklist and the rating scale.

2.5 List of resources

A document detailing the resources needed for the station should be prepared, in addition to the suitable examiner and the appropriate item of examination. The list shall be complete and include furnishings to be included in the room, such as desk and two chairs, X-ray box, wall mounted BP machine and auroscope and fundoscope. Where pre-prepared packs are required the contents shall also be specified. It shall

be noted that all such equipment must be made available preferably at least two weeks before the scheduled date of the examination.

2.6 Stations refinement, calibration and testing - formative OSCEs

In the meeting set for the review of the stations, the station constructor will present and explain the components of the station. At the end of the meeting participants decide to accept or reject the stations presented, and where needed after making any necessary modifications, which may mean asking the designer to make major revisions. New stations should be scheduled for testing and calibration where examiners make sure that the task is objective and doable and that the time suggested is appropriate. If a clinical sign needs to be demonstrated by the candidates with a non-standardized patient, the examiners must make sure that the sign is clear and can be detected by the examinees before the OSCE starts, and this can only be done by the examiner examining the non-standardized patient himself, and not during the OSCE itself.

2.7 Preparing the examinee's exam book

An OSCE book is desirable, but might not be favoured by every OSCE Committee. Not everything goes into the OSCE book, for example the Observer Examiner station mark sheets are retained at the station. Exam books were found to be very useful for the candidates especially when there are stressed.

It is recommended to include the following in the book:

1. The map of the venue.
2. The instructions and regulations for the exam (for example, five minutes reading time are allowed at the beginning of the exam).
3. The scenario and instructions for every station (active and static).
4. Marks awarded for each component of the station.
5. Answer sheets, in most cases with the questions, for example, X-ray and data interpretation. Usually with ruled lines and always with available marks stated, and includes only those marks/decisions (see above) that are allowed
6. Available resources. It needs always to be stated what the candidate should expect in general at each station, for example, a simulated patient or simulator, equipment (but not the specifics), an Observer Examiner (if not, would there be video observation and recording). The Observer Examiner's role needs also to

be clarified, "he/she will observe and rate the performance and ask two questions" - which should always be stated in writing in the Observer Examiner's instruction sheet.

2.8 Recommended standards and regulations

1. Every station must correspond to a specific item (cell) in the test blueprint.
2. Where the programme curriculum is available the construct must be clearly stated, including what level of mastery, and whether with a simulated patient or non-standardised patient, mannequin or simulation equipment.
3. Where possible, non-medical expert competencies stated by CanMeds should be given a place in the assessment being made, for example before a procedural demonstration the candidate shall explain to the examiner, as though a patient, exactly what they are going to do.
4. Every station must have an explicit marking schedule, and all stations must have one set of total marks, for example ten or twenty.
5. The following materials must be available for every station:
    a. The station title is implied through its classification, and must have at least two parameters
    b. Examiners' instructions must be available, including the construct being tested with an extract from the relevant part of the curriculum, as well as detailed instructions for the allocation of marks
    c. If relevant, there must be simulated patient/non-standardised patient instructions, which may include a simplified marking schedule
    d. An equipment list, including disposables and other materials must be provided for every station, this list shall state for a communication station for example, a desk and two chairs
    e. An instruction sheet that must be laminated and placed at the station as well as in the OSCE book, these instructions must include the distribution of marks against the tasks requested
    f. Computer based stations are legitimate provided they fit the test blueprint and are aligned with the curriculum, see above
    g. Stations which can be delivered in another format should, however, be avoided
6. OSCE stations must focus on appropriate psychomotor skills at the 'shows how' level of the Miller's pyramid avoiding items that can be included in tests of knowledge like MCQs, Spotter exams, bearing in mind that multiple-choice question tests for postgraduate examining will in the future be computer delivered which is an ideal medium for item enhancement with multimedia.

7. As much as possible, OSCE stations should be of the integrated type that assess soft skills such as communication, ethics, attitudes, and professionalism as well as the clinical task applicable to that station.
8. It is recommended that the number of OSCE stations should be between 10 - 15 stations (rest stations not included). The exact number can be decided after piloting.
9. Banking of OSCE stations is encouraged after applying all these guidelines with the feedback from candidates and examiners, so that the relevant station could be modified and re-tested.
10. Before adding a station to the relevant specialist stations bank, it is recommended that the quality of the stations be checked and approved using a quality control process, maybe using VPIs.
11. All newly designed stations shall preferably be calibrated and tested in a formative test before banking,

2.9 Question banking and station materials

The Saudi Commission for Health Specialties is developing an OSCE station banking software for storing and retrieving stations that have been created for Saudi Board Examinations. To make this possible it is necessary when creating a station that all the resource materials needed shall be created in separate MS Word files. The materials will differ depending on what type of station is needed, and below are examples of different types of stations (by no means exhaustive and often specialty specific):

1. History taking from a non-standardised patient (for example, a patient with systemic lupus erythematosis).
2. History taking from a simulated patient (another example, a case of tertiary parathyroid disease).
3. Counseling a simulated patient, examples might include cancer, pre-operatively, smoking, diabetes.
4. Information giving to a simulated patient, for example the implications of a diagnosis just made, such as Sjögren's, Barth syndrome or Batten's disease.
5. Information about drugs prescribed to a simulated a patient (SP).
6. Physical examinations, including:
    a. Breast exam, usually using a male simulated patient
    b. Thyroid exam (SP)
    c. Abdominal exam (SP)

   d. Neurological exam (SP)
   e. Specific tests (SP)
   f. Allen's test
   g. Trendelenburg's test
   h. Examinations of the pulses
   i. Examination of a groin swelling
   j. Assessment of varicose veins
   k. Testing hearing
   l. Fundoscopy
7. Surgical procedures, for example colon anastomosis on a jig
8. Measuring the JVP using an adjustable model
9. Application of traction on an SP
10. Laparoscopy
11. A case of mitral valve disease with essential investigations (echocardiogram, ECG and chest X-ray)
12. A series of high fidelity breast simulations and corresponding mammograms
13. Determination of Glasgow Coma Score on an SP and the management
14. Demonstration of sterile technique and male catherisation with mannequin

The resource materials must include some or all of the following:

1. Candidate instructions.
2. Examiner's instructions, including grading.
3. Simulated patient instructions including history to be given if relevant (shall be comprehensive).
4. Instructions for a non-standardised patient.
5. Equipment list, including disposables and sterile packs, X-rays, special equipment such as surgical jigs and charts.
6. Special requirements, for example a technician to replenish materials or repack equipment.
7. A scoring form

Chapter Three: Exam Construction

3.1 The Test Blueprint

Developing a test blueprint is the process by which test content is carefully planned and then aligned against curricular learning objectives. Test blueprints are important in any assessment to ensure content validity; the most important concept in any assessment system. A test blueprint ensures that the test measures the intended learning outcomes. It is recognised that no test can measure everything in a course/programme or whole curriculum. The test should sample relevant and representative course content using a test blueprint or a specification table.

A test blueprint or specification table could be prepared by plotting the program contents against the program objectives to produce a grid or table, then putting the components of the program and their weights in the rows and the tasks or competencies in the columns. The the number of stations for each component is calculated and distributed across the competencies which are most appropriate for the component.

Preparing a Test Blueprint for an OSCE

This task is the first duty of the OSCE committee. Validity is a requirement of every evaluation and implies that candidates achieving the acceptable performance level have acquired the level of competence set out in the learning objectives. Typically, the type of validity that relates to measurements of academic achievement is content validity (Hopkins 1998). Evaluation of content is considered to be valid when it is congruent (aligned) with the objectives and learning experiences, and congruence between these pillars of education can be facilitated by using an evaluation blueprint (Bordage et al. 1995; Bridge et al. 2003).

Blueprinting is therefore, a process by which educators ensure that all important components are proportionately assessed.

There should be distribution of cases whereby all competencies are mapped. The focus must be on competencies rather than selecting cases based on their availability in an OSCE assessment bank. Another point to be considered is the weight of each item in the assessment blueprint.

Based on the classification proposed earlier, to prepare a test blueprint for an OSCE a grid can look like the one depicted below, Table **3.1**. The components of the educational units are then listed in column 2 and their weight in column 3. After the

examiners agree on the number of stations in the exam, the numbers for each component are calculated according to its relative weight. The stations for each component are then distributed to the most important competencies for that component. If not all components will be included in the exam, list only the components to be included and recalculate the relative weight.

Table 3.1. An example test blueprint

No.	Program Component	No of stations	Domains of clinical competence						
			Communication		Examination		Procedures		
			HT	OC	PE	VSI	TP	DP	IATF
			A	B	C	D	E	F	G
1	Cardiology	3	1		1		1		
2	Respiratory	1							1
3	Endocrine	2		1	1				
4	Neurology	3			1	1			1
5	Oncology	2	1	1					
6	Gastrointestinal	3	1		2				
7	Nephrology	2	1						1
8	Dermatology	1				1			
9	Rheumatology	1	1						
10	Haematology	1							1
11	Other	1		1					
Total		20							

- **Communication: HT=**Focused History Taking, **OC=**other communication.
- **Physical Examination: PE =** Physical examination, **VSI=**Virtual Sign Identification.
- **Practical Procedures: DP=**Diagnostic Procedure, **TP=**Therapeutic Procedure, **IATF=**Identification of Abnormal Test Finding.

3.2 Converting the blueprint into a test map

The next step in the preparation of an examination is the conversion of the blueprint into a test map. This can be achieved by the following steps:

i. Draw a table as shown in Table **3.2** (components and tasks/skills)

ii. Identify the conditions/problems which were agreed by the OSCE exam committee members for the different systems; for example; chest pain (CVS), MI, cough (CVS), haemoptysis (RS), joint pain (MS), headache (NS) and so on.
iii. Agree on the weight given to each problem and possibly more competencies to be tested for any specific problem.
iv. Transfer these problems/diseases in the appropriate rows according to their related system and competencies tested.
v. Be sure that all or a majority of cells are filled with the competencies to be tested, Table **3.2**.

When an OSCE blueprint is ready, a station list can be prepared through the identification of the clinical conditions and tasks for each component of the educational unit. The cells in the blueprint grid are named according to the cross-section of the column and row in which they lie. The committee then decides on a clinical condition and task for each selected cells. If a cell in the blueprint contains more than one station, the conditions are numbered 1, 2. If they are two stations in cell 6C, the stations would be 6C-1 and 6C-2, see Table 3.1. A list of stations is, therefore, created. If there are stations in cells 1A, 1D, 1F, 1H, 2B, 2F, 2I, 3E, a station list might look like the following:

1A Man with a cardiovascular problem (history taking - the technique on a SP)
1D Person with a heart problem (examination of the precordium - the technique on a NSP)
1H Man with cardiac arrest (CPR - a therapeutic procedure)
2B Young man with a history of pneumonia (focused history with SP)
2F Video of a distressed asthmatic with audible wheezes (virtual identification of the signs)
2I Chest X-ray of a man with a pleural effusion (Identification of abnormal test findings)
3E An SP with hepatitis and tender hepatomegaly (physical examination - sign elicitation), and so on.

The committee must ensure that the test is balanced as far as the three major competence categories and the components of the educational units and these in turn depend on the nature of the unit. A unit in which procedures are most important, more cases addressing the competencies are expected and so on.

The following template can be used for the preparation of the station list.

Table 3.2

Blueprint Cell		Item title (condition)	Competence tested	Proposed test item
A	1	Myocardial infarction	HT	Simulated pt
C	1	Hypertension and CHD	PE (Auscultation)	Mannequin or pt
E	1	Cardiac arrest	TP (CPR)	Mannequin
G	2	Pneumonia	IATF	Chest x-ray
B	3	Type II Diabetes Mellitus	OC (counselling)	Simulated pt
C	3	Thyrotoxic Goitre	PE (Head and Neck)	Real patient
C	4	Hemiplegia	PE (neurologic exam)	Real patient
G	4	A bleeding AVM	IATF	MRA
A	5	Carcinoma of the stomach	HT	Simulated pt
B	5	Breast Cancer	OC (breaking bad news)	Simulated pt
…	…	…	….	…..

3.3 Deciding on types of stations (manned or unmanned, dynamic or otherwise, linking or duplex)

The main purpose of a Postgraduate OSCE is to test clinical competencies at the 'show' level of the Miller pyramid. Therefore, when preparing the list of OSCE stations, after developing the test blueprint, only those stations should be included which really test the 'show' level of the candidates. The following are some of the station formats:

i. **Integrated stations.**
   These are stations in which two of more competencies are tested for example history and examination, history and data interpretation, history and communication (counselling, getting informed consent, breaking bad news, and so on).

ii. **Linking or duplex stations**
   Usually, the time spent for an OSCE station is 5 - 15 minutes (undergraduate level, 5 - 7 minutes; and postgraduate level, 8 - 15 minutes). Duplex stations allow two related tasks to be tested, for example, at the first station a history is

taken from a simulated patient, and at the following station lab or imaging results of the case are displayed for identification of abnormalities and interpretation of data, or a questionnaire is used related to the history just taken.

Alternatively, some skills may need more than the prescribed time and two parallel stations are placed in the OSCE circuit and candidates are fed first to one and then the other thereby allowing double the time at the station.

3.4 Preparing the exam materials (test book, scoring forms, instructions)

Preparation of examination materials for OSCE stations takes time and effort. This preparation should use a team approach to take on the different responsibilities and tasks needed.

Below is a simplified step-by-step approach for preparing OSCE materials:

1. If an OSCE blueprint is not available, the OSCE committee will start by preparing one following the steps described above (3.1.)
2. If an OSCE station bank is available and the stations are on the committee's station list, the material needed is *pulled out* of the bank. If no bank is available the task should be assigned to groups of station designers (constructors).
3. An examinees sheet is prepared that includes the station profile, the opening statement (scenario or preface) and the candidate instructions.
4. An examiner's sheet should contain the station profile and the examiner's instructions accompanied by a scoring form (checklist or rating scales).
5. Instructions to patients are pulled off the bank or freshly prepared and used for the training of patients (real or simulated)
6. Tasks and instructions for the support team are prepared describing the members of the team and their responsibilities.
7. Station labels, instructions and directions are prepared
8. An exam check list with deadlines is compiled detailing what needs to be prepared so that the committee will not forget any issue or item

3.5 Recommended standards and regulations

1. The OSCE prepared must fulfil the 'Utility Criteria' which include validity, reliability, educational impact, cost-effectiveness, and acceptability. To fulfil

these criteria robust and high quality standards should be applied. For this purpose a quality assurance process also has to be agreed-upon.

2. To ensure content validity, it is recommended that each board creates a master blueprint which specifies the different assessment activities and methods used to assess all the content and the learning domains/outcomes derived from the curriculum of the specialty. The OSCE might be one of several methods used for Clinical/Practical (Saudi Board Part 1) or final assessment (Saudi Board Part II).
3. An OSCE blueprint can then be developed from the master exam blueprint
4. A test map of the venue to be used and a detailed station list should then be developed from the blueprint together with an exam follow-up checklist for organizational purposes
5. Preferably stations from the station bank should be used in high stakes OSCEs
6. Since these OSCEs are for postgraduate candidates, it is recommended that rating scales be used as much as possible.

Chapter Four: Examination Logistics

4.1 Assigning roles and responsibilities (exam coordinator, examiners, support staff)

The preparation of an OSCE is different from preparing knowledge based exams with multiple-choice or essay questions. The success of an OSCE largely depends upon thorough planning and efficient organization. Without comprehensive planning and organization, the OSCE will be flawed and its implementation risks being unsuccessful. Though the organizational structure of running an OSCE will depend on local circumstances, a majority of the components are common to all settings. The various components involved in OSCE planning and management and the individual roles involved will be described:

i. OSCE Coordinator

While decisions such as the content and scope of the examination require input from the Exam committee, there must be a single committee member to coordinate the planning and implementation of the exam. The coordinator is responsible for overseeing the development, organization, administration, and grading of the examination. In the case of multisite examinations, there should also be local coordinators available for each individual site. The coordinator should be identified as early as possible in the planning of an OSCE. It is essential that this person should have had previous experience managing an OSCE or at least being part of an OSCE exam committee, especially if the OSCE is being administered for the first time. The OSCE coordinator should have strong motivation, be well connected to resources including in-house and outside testing facilities, and be able to communicate well and create team spirit.

ii. OSCE Committee

This committee is responsible for overseeing the whole examination process from planning to the final publication of the score report. The committee should consist of the coordinator, a small number of senior specialists familiar with the curriculum and desired standards, a simulated patient coordinator/ trainer, a supervisor for the support staff and an educationist familiar with performance based testing. This committee will approve the test blueprint, the selection of station developers, appointment of examiners, decisions about pass and fail criteria.

Since the OSCE is a resource demanding exam, the committee should hold its first meeting at least three months before the date of the exam and assign tasks with

deadlines to each individual involved in the exam. The committee should then hold frequent follow-up meetings to determine the progress of assigned tasks and troubleshooting issues without delay. The more members who have had previous experience with running an OSCE the more likely the final product will be successful.

iii. Station Developers

Once the test blueprint is approved, the task of developing stations can begin by assigning specific stations according to the test blueprint. The appointed station developers must have good clinical experience, should be familiar with the curriculum or training programme and published standards, and be a good team member who does not take offence when the committee offers suggestions for station revisions. All the stations developed must be submitted to a central office, preferably to the Commission's Department of Medical Education and Postgraduate Studies for transforming into a unified format and undergoing a technical review. After the necessary technical review, all the formatted and edited stations should be presented to the OSCE committee for scientific review and to determine the compliance of the station construction with the blueprint and measurement validity

If the OSCE is going to be held for first time a workshop may be very useful to help the examiners develop sound and valid stations in the prescribed format.

iv. Support Staff

In addition to the faculty and station developers, recruitment of support staff is equally important for a successful OSCE. This should comprise of at least one secretary and a senior person experienced in liaising with logistic and support services people to resolve the physical issues. The following are the major responsibilities of the support staff:

- Photocopying
- Preparing and distributing materials
- Distributing materials
- Setting up individual stations
- Attending the needs of examinees, examiners and simulated patients
- Developing the OSCE map
- Setting up the bell system
- Developing and placing the number and arrow signage at appropriate places
- Arrangement of required material and equipment
- Maintaining a central store for necessary material, equipment and their replacements
- Quarantine arrangements for examinees waiting for their exam

- Collecting answer sheets from every station and examiner. In the case where an OSCE book has been used, separating the relevant sheets with several colleagues to deliver for marking in appropriately prepared envelopes
- Entering scores in a spread sheet
- Catering for all personnel involved during the day of exam

v. Trainer

Simulated patients may be required to perform a variety of tasks, for some of which they may need training. Sessions must be arranged at which SPs can be trained by an appropriate and experienced person. The trainer must understand the roles of simulated patients, examiners and specific case requirements. A trainer must also have good teaching skills and be capable of providing constructive feedback. The appropriate person may be a skills trainer from a skills centre but it is always preferable that the trainer would be the same clinician who developed the station. If the trainer is not the station developer then he must meet the developer of each station to discuss the patient profiles and obtain a detailed understanding of the problem to be presented. The trainer should also be present on the day of exam to deal with any simulated patient problem.

vi. Examiners

Examiners are required for the stations where clinical skills such as physical examination, history taking, interviewing or counselling will be assessed.

The examiner must be clear about the objectives and standards, fair and free of any personnel style and biases and be able to provide effective and constructive feedback, if required. Do not use external examiners as examiners; they are far more valuable giving feedback on all stations.

vii. Time keeper (support staff)

The role of the time keeper is to maintain the OSCE schedule using bell rings at very precise intervals. The time keeper should remain focused and not be distracted. This task is far more important than the recognition usually given to it. It is also the sort of task which needs a rather obsessional type of person to carry out the duty successfully.

viii. Monitor (senior member of the OSCE Committee)

The role of the monitor is also very important because he is responsible to directing rotation flow, identifying and solving issues that may arise with examiners or simulated patients as well as unforeseen shortages of necessary materials. This person would always have the authority to stop the OSCE if the necessity arose, and

planned bell rings should be arranged in that event to stop and then later to re-start the OSCE.

4.2 Simulated patients / standardized patients

Standardized patients who are the real patients could be used in the OSCE but it may be difficult to find standardized patients with similar clinical features for more than one OSCE circuits. Using non-standardised patients can be more problematic than using simulated patients. So, most of the time the use of simulated patients is preferable because of their availability, more easily reproduce the same case for all examinees, and able to give feedback (if necessary) as well as providing opportunities for sensitive consultations. It has been shown that a well-trained simulated patient can be used not only to present the patient in a consistent and reliable manner, but also to evaluate the clinical skills of the examinee. The simulated patient should be committed to repeatedly portray a clinical condition with standardization, confidently, in a comfortable manner and take educational responsibility if required.

Being repeatedly interviewed or examined is always tiring and sometimes uncomfortable. Therefore, it may be desirable to train two simulated patients for such a station, particularly if the exam is for a longer duration or if the station is demanding.

Simulated patients can be volunteers or professional actors. The volunteers may include colleagues, friends, students or hospital staff. Whoever plays the role of a simulated patient, the payment policy must be clear. It is also important that all payments are made in a timely manner.

4.3 Orientation and debriefing of the examinees, patients and observers/examiners

Orientation of examiners, examinees and patients is essential. It is preferable to send out briefing notes and details of the stations to Examiners and simulated or non-standardised patients at least two weeks before the OSCE. Prior dissemination of detailed information relating to the exam greatly facilitates the process. Examiners, examinees and patients should be asked to arrive at the examination site at least 30-45 minutes prior to the examination. Usually the only specific briefing needed on the day of the OSCE is for examinees, although someone must be available to guide examiners and patients to their stations and ensure that they know where to find any essential materials, for example, the marking sheets and clip board.

It is also valuable to have a debriefing meeting of the organizers shortly after the examination to review all aspects of content and organization. More value will be added if feedback has been obtained from examiners, simulated patients and a representative of the examinees. This feedback provides useful information about the performance of each station and can assist decisions made to modify and improve stations before being banked for future use. A written summary to improve administrative and organizational aspects will be a valuable reminder when the time arrives for the next OSCE.

4.4 Choosing and preparing the venue including the use of video technology

If a dedicated OSCE centre or custom built skills lab are not available to hold the OSCE in, a good alternative venue is a clinical environment. An outpatient clinic or an empty ward will work well. The main disadvantage to using outpatient clinics is the need to schedule the examination so that it does not conflict with normal patient care activities. Therefore, most examinations that utilize clinical facilities may need to take place during the evening hours or on weekends, alternatively be booked a long time in advance (up to one year). Clinical settings may need to be re-arranged to provide a well-functioning venue, in particular the layout should be planned to allow the smooth flow of examinees from station to station. When planning an OSCE, the testing area should be mapped, station placement and type (for example, unmanned, manned and with patients) noted and flow patterns shown. The layout should be such that the bell(s) used to indicate the time to change stations can be heard throughout the examination area. The best choice of bell are those frequently used for fire alarms. Use of video technology in OSCEs is increasing. To save time and money, some of the stations involve videotaped material. Videotapes can also be used to enhance the quality of training simulated patients. Some of the venues also offer the facility of video recording to capture the encounters at each station (the Saudi Commission has three such facilities). The venue also contains a control room where real time video feed from each station can be monitored remotely on display monitors and scored by examiners.

4.5 Preparing the exam materials (signs, bell system, exam items, patients, catering)

This preparation should be done at least one day prior to the exam: preparing the physical facilities in which the OSCE is to be conducted plays a vital role in the

success of the exam. This will involve the arrangement of furniture, screening of areas where patient are to be examined, setting up and equipping the stations, labelling stations, posting signage and arrows, and ensuring that the bell system is working. To accommodate the number of candidates a morning and afternoon session may be necessary and so catering arrangements must be provided for examiners and simulated patients. Where groups of examinees are coming for a later session (for example afternoon), quarantine arrangements must be arranged for those who have just finished their OSCE, and in which case some refreshment is always welcome.

4.6 Running the OSCE

On the day of exam, the OSCE coordinator and support staff should arrive early. Simulated patients should arrive next and where non-standardised patients are required they come last accompanied by nursing staff. Examiners should arrive at least 45 minutes before the start time to familiarise themselves with the details of the station and also with their patients. It is important that examiners are briefed some days before the OSCE itself so as to leave enough time to answer all their questions. Candidates should arrive at the venue at least 30 minutes prior to the start time (in medical schools a lecture theatre is the ideal collection point). The candidates should be given a final briefing, including details of any quarantine arrangements, their starting positions and the movement flow in the OSCE circuit, and finally some words of encouragement.

When everyone and everything is ready, the bell should ring to commence the exam. The candidates, examiners and all staff involved shall be fully informed of the bell signal formula (start, stop, re-start and so on). During each change over, organizers must ensure the exam is running smoothly. For high stakes summative exams maintaining security is important. The examinees and staff running the OSCE must be aware of the regulations, which shall be common for all postgraduate examinations, and issued by the Saudi Commission.

4.7 Marking

All mark sheets are collected from each station and checked for candidate identification, name and / or number. It is advisable to enter the marks into a data sheet (Microsoft Excel) as soon as possible after each session so the scores can be compiled shortly after the last OSCE session is completed. Adequate time should be allowed to process the marks and the data manager should be supported by sufficient staff to ensure no candidate score is omitted.

4.8 Recommended standards and guidelines for the delivery of exams

1. An OSCE committee should be appointed with an exam coordinator to supervise the delivery process helped by a support team.
2. A suitable exam venue should be chosen, prepared and tested well ahead of the exam day with sufficient space for patients, examiners at their stations and briefing and quarantine rooms for candidates.
3. The selection of examiners should be based on specific criteria, including training and contribution/participation in previous OSCEs.
4. To reduce variability and improve inter-rater reliability, workshops/training courses must be organized for the purpose of training and orientation of examiners and simulated patients.
5. The entry and collation of candidate marks should only be done by support team staff as soon as possible after each OSCE session.
6. The evaluation of students at each station could be done by examiners physically present at the station or through a closed-circuit video system.
7. Instructions, answers and reflection could be done on papers posted and located at each station or preferably in an examinee's OSCE book distributed to candidates at the beginning of the exam.
8. Debriefing of examiners, patients, candidate representatives and support staff should be done as soon after the day of the OSCE as is convenient.
9. Some of the stations will need to be videoed for standardization purposes and to give further feedback to examiners.
10. A workshop has to be delivered by the Saudi Commission to train examiners and station developers.
11. *Post hoc* examination analysis will be carried out by the Assessment Division of the Department of Medical Education and Postgraduate Studies. Recommendations about station numbers shall be made where the exam reliability falls substantially below 0.80.

Glossary

Assessment

A method to determine level of accomplishment for a specific purpose, for example promotion. Assessments may be formative, and be part of the learning process and characterized by feedback, or summative, in which case the results are used to make an important decision, selection, certification, promotion or graduation. Assessments may also be characterized as criterion referenced with specific competencies aligned with the assessment items, or norm referenced when the candidates are ranked in order with a cut off set at some point in the ranking, using items with a difficulty often beyond the period of study.

Checklist Evaluation

An assessment method based on an expert consensus making a list of behaviours associated with a specific task. For each behaviour a judgment is then made using a rating scale, or in some cases a dichotomous scoring method (one or zero depending if the behaviour was observed or not).

Clinical Competence

The ability to provide total care of patients by those charged with their care. Clinical competence may be further defined according to level of training and experience. Clinical competence is not usually considered completely measurable, rather only certain abilities such as knowledge or psychomotor skills.

Communication skills

The proficiency of verbal and non-verbal exchange of information, ideas and decisions. The use of language as means to obtain information, convey empathy and provide support for patients and their relatives.

Competence

The possession of sufficient amounts of knowledge, skills and appropriate attitudes needed for accomplishment of specific tasks. Once again not all these attributes are amenable to testing, becoming less so the higher the professional hierarchy.

Educational or Instructional Objectives

These describe what learners are expected to master, perform or demonstrate with the underlying knowledge, skills or attitudes.

Evaluation

A process of making comparative judgments about learners, trainees or a training program. Evaluation may be formative or summative. Formative evaluation has two

components (a) diagnostic and (b) feedback. Summative evaluation measures whether specific objectives have been met.

Global Rating and Global Rating Scales

Two methods for grading performance either as a holistic judgment (global rating) or against a list of criteria related to the scale provided. In essence both these methods allow for expert judgments.

Performance

This is an observed activity and may be in a real (a doctor in his outpatients) or simulated situation (a medical student in an OSCE).

Reliability

The reliability of a test is the measure of its score precision, consistency and reproducibility. It is given as a coefficient when 1.0 means absolute reliability and 0.0 none. A knowledge based test for a high stakes examination a coefficient above 0.86 is desirable.

Simulated Patient (SP) and Non-Standardised Patient (NSP)

An SP is a fit and well person who is playing the role of a patient with a specific medical problem. SPs are usually used in a formal assessment setting, for example an OSCE. An NSP can also be used in a similar formal assessment setting.

Simulations

This is a piece of equipment or model that mimics real life-like situations, an example might be the use of a mannequin for resuscitation purposes.

Validity

Validity is the degree to which a test measures what it is supposed to measure. Four other forms of validity have been described:

1. Content Validity: A test has content validity when what is being tested reflects what has been studied (or taught) or delivered by the curriculum.
2. Face validity: The perceived relationship of an item by a candidate with the intention being tested.
3. Concurrent Validity: The degree to which the test gives the same information as another form of test measuring the same thing.
4. Predictive Validity: The extent to which a test predicts future outcomes.

The Saudi Commission for Health Specialties
الهيئة السعودية للتخصصات الصحية

www.scfhs.org

ISBN: 978-9960-9832-8-8