# Objective Structured Clinical Examination (OSCE) Revisited

**PIYUSH GUPTA, POOJA DEWAN AND TEJINDER SINGH***

*From the Department of Pediatrics, University College of Medical Sciences, Delhi, and *Department of Pediatrics, Christian Medical College, Ludhiana, India.*

*Correspondence to: Dr Piyush Gupta, Block R 6 A, Dilshad Garden, Delhi 110 095. prof.piyush.gupta@gmail.com*

Objective structured clinical examination (OSCE) was introduced in 1975 as a standardized tool for objectively assessing clinical competencies - including history-taking, physical examination, communication skills, data interpretation etc. It consists of a circuit of stations connected in series, with each station devoted to assessment of a particular competency using pre-determined guidelines or checklists. OSCE has been used as a tool for both formative and summative evaluation of medical graduate and postgraduate students across the globe. The use of OSCE for formative assessment has great potential as the learners can gain insights into the elements making up clinical competencies as well as feedback on personal strengths and weaknesses. However, the success of OSCE is dependent on adequacy of resources, including the number of stations, construction of stations, method of scoring (checklists and/or global scoring), the number of students assessed, and adequate time and money. Lately, OSCE has drawn some criticism for its lack of validity, feasibility, practicality, and objectivity. There is evidence to show that many OSCEs may be too short to achieve reliable results. There are also currently no clear cut standards set for passing an OSCE. It is perceived that OSCEs test the student's knowledge and skills in a compartmentalized fashion, rather than looking at the patient as a whole. This article focuses on the issues of validity, objectivity, reliability, and standard setting of OSCE. Presently, the Indian experiences with OSCE are limited and there is a need to sensitise the Indian faculty and students. A cautious approach is desired before it is considered as a supplementary tool to other methods of assessment for the summative examinations in Indian settings.

**Key words**: *Assessment, Clinical, Competency, India, OSCE, Reliability, Validity.*

That '*learning is driven by assessment*' is a well known fact. This is also referred to as the '*steering effect of examinations*'. To foster actual learning, assessment should be educative and formative. Medical education aims at the production of competent doctors with sound clinical skills. Competency encompasses six inter-related domains as developed by Accreditation Council for Graduate Medical Education (ACGME): knowledge, patient care, professionalism, communication and interpersonal skills, practice based learning and improvement, and systems based practice(1). Epstein and Hundert have defined competence of a physician as "the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values and reflection in daily practice for the benefit of the individuals and the community being served"(2). The community needs to be protected from incompetent physicians; and thus there is a need for summative component in the assessment of medical graduates.

## LOOKING BEYOND THE TRADITIONAL TOOLS

The traditional tools for assessment of medical students have mainly consisted of written exams (essay type, multiple choice, and short-answer type questions), bedside viva and clinical case presentation. These have focussed on the "*knows*" and "*knows how*" aspects, i.e., the focus has been on the base of the '*Miller's pyramid of competence*' (***Fig.*1**). These methods of assessment however have drawn a lot of criticism over the years because of their inability to evaluate the top levels of the pyramid of competency in a valid and reliable manner. The following flaws were realised:

**WHAT IS AN OSCE?**

**O**bjective • **S**tructured• **C**linical• **E**xamination

1. Ensures evaluation of set of predetermined clinical competencies.
2. Each clinical competency is broken down into smaller components; e.g., taking history, performing examination, interpreting investigations, communicating, etc.
3. Each component is assessed in turn and marks are allotted according to predetermined checklists.

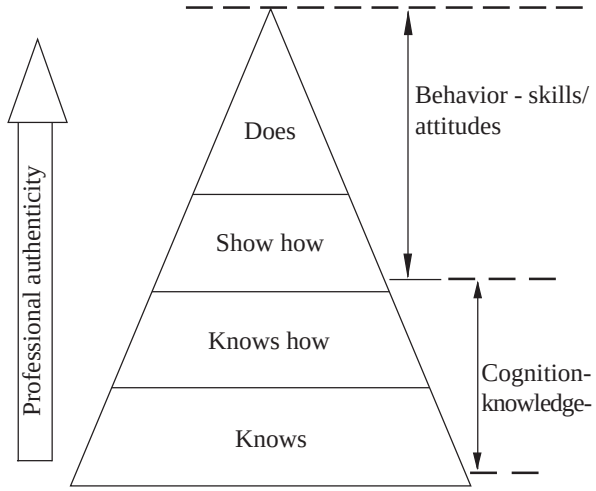


**FIG.1** *Miller's pyramid.*

- They test only the factual knowledge and problem-solving skills of students, which may be appropriate only in the early stages of medical curriculum. These methods do not evaluate the clinical competence of students. Important aspects like performing a particular physical examination (shows how), clinical maneuver, and communication-skills are not tested. Only the end result is tested and not the process of arriving at a result.
- The students are tested on different patients (patient variability). Each student is adjudged by only one or two examiners, thereby a scope for marked variation in the marking by different examiners (examiner variability). These factors increase the subjectivity of marking (lack of reliability).
- There is often a lack of clarity on what is actually being tested (lack of validity). Assessment is usually global and not competency based.
- Students are not examined systematically on core procedures.
- There is no systematic feedback from the students and teachers.

To obviate the drawbacks of conventional clinical evaluation, objective structured clinical examination (OSCE) was first introduced by Harden in 1975, as a more objective, valid, and reliable tool of assessment(3). In an ideal OSCE, all domains of competencies are tested, specially the process part; the examination is organized to examine all students on identical content by the same examiners using predetermined guidelines; and a systematic feedback is obtained from both students and the teachers. OSCE is meant to test the '*shows how*' level of the Miller's pyramid(4).

## CONTENT AND PROCESS OF OSCE

OSCE consists of a circuit of stations which are usually connected in series (***Fig.* 2**). Each station is devoted to evaluation of one particular competency. The student is asked to perform a particular task at each station. These stations assess practical, communication, technical, and data interpretation skills and there is a predetermined decision on the competencies to be tested. Students rotate around the complete circuit of stations, and keep on performing the tasks at each of the stations. All students move

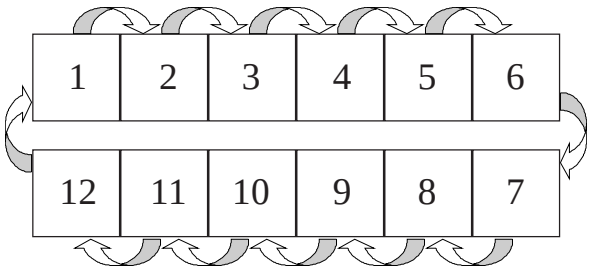


**FIG.2** *OSCE consists of a circuit of stations which are usually connected in series.*

from one station to another in the same sequence. The performance of a student is evaluated independently on each station, using a standardized checklist. Thus, all students are presented with the same test; and are assessed by the same or equivalent examiners. Students are marked objectively on the checklist(5) by the examiner.

***Types of OSCE stations:*** The stations are categorized as '*procedure station*' or '*question station*'. Procedure stations are observed by the examiner while question stations are unobserved (only a written answer is desired). Student performance on a Procedure station is observed and marked there and then only while the Question stations can be evaluated later. The details of these stations along with specific examples have been described previouly(5). Procedure station and a question station can also be used together. In the original description of OSCE by Harden, every Procedure station was followed by a Question station. Students are given a task to perform in Station 1 (which is observed and assesses the process of performing the task) and the questions are presented later (in Station 2). Questions in station 2 are related to station 1 only. This has two advantages: (*a*) different domains of learning can be assessed by them; and (*b*) the effect of cueing is minimized. It is also advisable to incorporate a *rest station* for every 30-40 minutes into the exam, to give a break to the students, the observers and the patients. They also allow time to substitute patients at a clinical station, or to complete the written left over task from the previous stations.

***OSCE setup*:** The number of stations can vary from 12 to 30 though usually 20 stations suffice(1). The usual time allotted is 5 minutes for each station; ACGME however recommends station duration of 10-15 minutes. Giving more time per station allows more competencies to be tested in relation to the given task. All students begin simultaneously. The number of students appearing in the exam should not exceed the number of stations. In case, the number of students is more, one or more parallel sessions can be organized, subject to availability of space, examiners and patients. If facilities do not permit this, then two sessions can be planned. All students should commence the examination from a procedure station. The entire exam is usually completed within 60-150 minutes. Details of microplanning of OSCE have been described earlier(6).

**Blueprinting: Preparing the Stations**

Once the consensus is reached on the number and type of stations to be included, the next task is to formulate the questions, model keys, and checklists for each station. When planning an OSCE, the learning objectives of the course and the students' level of learning need to be kept in mind. The test content need to be carefully planned against the learning objectives - this is referred to as "blueprinting"(7). Blueprinting ensures a representative sample of what the student is expected to have achieved. Blueprinting in practice consists of preparing a two-dimensional matrix: one axis represents the competencies to be tested (for example: history taking, clinical examination, counseling, procedure) and the second axis represents the system or problems on which these competencies are to be shown (for example: cardiovascular system, nutritional assessment, managing cardiac arrest, etc.)(8).

Blueprinting is essential for building a higher *construct validity* of OSCE by defining the problems which the student will encounter and the tasks within the problem which he is expected to perform. By laying down the competencies to be tested in a grid, the correct balance between different domains of skill to be tested can be obtained (***Fig.* 2**).

Clinical competencies (including psychomotor skills and certain affective domains) should be primarily identified and included in the OSCE setup. OSCE can test a wide range of skills ranging from data gathering to problem solving(4). Although it can be used for this purpose, OSCE is not very suited for evaluating the cognitive domain of learning, and certain other behaviors like work ethics, professional conduct, and team-work skills. For these objectives, it is appropriate to use other modes of assessment. Feasibility of the task is equally important. Real patients are more suited to assessing the learner's examination skills while simulated patients are more suited to evaluate the communication skills of the learner.

| | History | Examination | Procedure/data interpretation |
|---|---|---|---|
| **CVS** | Chest pain | Cardiovascular system | ECG interpretation; BP |
| **Chest** | Fast breathing and cough | Respiratory system | Chest physiotherapy/ Peak flow |
| **Abdomen** | Abdominal distension | Abdomen examination | Ascitic tap |
| **CNS** | Headache | Nervous system/ Eyes | Fundoscopy |
| **Cardiac arrest** | | | CPR |

FIG.3 *Grid showing the OSCE blueprint to assess final year medical students.*

Occassionally, faculty members designing OSCEs run out of ideas and end up preparing OSCEs assessing only the recall and by repeating an earlier OSCE station. The key to ensure that students' learning is not restricted to 5 or 6 stations commonly used in a subject, is to have a good blueprint of competencies to be tested and the stations should be rotated in different examinations.

## SETTING THE STANDARD

A major impediment in the success of OSCE remains '*setting the pass mark*'. The standards for passing OSCE can be either relative (based on norm-referencing) or absolute (based on criterion-referencing). Both have their own utility as well as merits and demerits.

### *Norm-referencing*

'*Angoff approach*' and '*borderline approach*' are commonly used to set relative standards for OSCE. In the former, expert judges determine pass marks based on their estimates of the probability that a borderline examinee will succeed on each item in a test(9). A major drawback of this method is that the overall performance of a candidate is not judged. Also the estimates are based keeping a hypothetical candidate in mind and therefore may be incorrect. This way, different pass marks will be set across different medical institutions(10). In addition, this is a time-consuming process and requires greater commitment from the examiners. A minimum of 10 judges are required to obtain reliable results (11).

The borderline approach (formulated by Medical Council of Canada)(12) is a simpler and more commonly accepted method for setting the pass marks. In this method, the expert judges score examinees at each station according to a standardized checklist and then give a *global rating* of each student's overall performance. The student can be rated as pass, borderline, fail, or above expected standard. The mean scores of examinees rated as borderline becomes the pass mark for the station and the sum of the means becomes the overall pass mark(13). To increase the reliability of this method all the expert judges should be subject experts and several examiners should examine at each station. The Otago study(14) showed that 6 examiners per station and 180 examinees are needed to produce valid and reliable pass marks. This method has gained wider acceptance because the pass marks set are actually an average of differences in opinion of examiners unlike the '*Angoff marks*' which are obtained by arguing out the differences in opinion of the examiners. Whatever method of standard setting it used, a fine tuning of the 'experts' is necessary so that they view the performance of the students appropriate to his level (e.g. undergraduate or postgraduate) and not from a specialist perspective.

Wass, *et al.*(7) state that "Norm-referencing is clearly unacceptable for clinical competency licensing tests, which aim to ensure that candidates are safe to practice. 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."

### *Criterion-referencing*

An absolute clear-cut minimum accepted cut-off is decided beforehand(15). For example, Medical Council of India (MCI) recommends 50% as the minimum pass marks for all summative examinations in medical specialities. National Board of Examination (NBE), India also accepts overall

50% marks as minimum acceptable for passing in OSCE examinations. A problem with using the overall pass mark as a benchmark for competence may not be acceptable as exceptional performance in a few stations would compensate for poor performance in other stations. It would be more appropriate to decide upon a minimum total score and a defined proportion of stations which the examinee must pass in order to pass the OSCE(13). Certain institutions also make it mandatory to pass the critical stations. However, it should be kept in mind that OSCE allows students to score much higher marks as compared to case presentation and adding the two to decide a pass percentage may be inappropriate. As a good practice, scores obtained at OSCE should be reported separately from the scores obtained at case presentations. Correlation between the two sets of scores is generally poor(16).

*Checklists vs Global Rating*

Checklists were designed and incorporated into OSCE to increase the objectivity and reliability of marking by different examiners. However, scoring against a checklist may not be as effective as it was thought to be(17). Evidence is accumulating that global rating by an experienced physician is as reliable as the standardised checklist. Regehr, *et al*.(18) compared the psychometric properties of checklists and global rating scales for assessing competencies on an OSCE format examination and concluded that "global rating scales scored by experts showed higher inter-station reliability, better construct validity, and better concurrent validity than did checklists. Further the presence of checklists did not improve the reliability or validity of the global rating scale over that of the global rating alone. These results suggest that global rating scales administered by experts are a more appropriate summative measure when assessing candidates on performance based assessment." Use of global ratings, however, mandates that only people with subject expertise can be used as examiners. However, there is still no consensus on the gold standard for the same. A balanced approach is suggested by Newble(8) wherein checklists may be used for practical and technical skills stations and global rating scales are employed for Stations pertaining to diagnosis, communication skills and diagnostic tasks. Another approach could be to use checklists during early part of clinical training and global ratings during final summative years.

*Example of a global rating scale for assessing communication skills*

*Task*: Counsel this 35 year old woman who is HIV positive about feeding her newborn baby.

The student is rated on a scale of 1-5. The examiner score sheet would read as follows:

| | | |
|---|---|---|
| 1. | Exceptional | ☐ |
| 2. | Good | ☐ |
| 3. | Average | ☐ |
| 4. | Borderline | ☐ |
| 5. | Poor/Fail | ☐ |

*Note: A checklist can be provided to assist the examiner in making his judgement of the student's performance, though no marks are decided for each item on the checklist. Using a checklist for a global rating can enhance the validity and reliability of OSCE.*

## THE CONCERNS

OSCE, now into 35th year of its existence, has had its share of bouquets and brickbats. Despite controversies, it has stood the test of the time and has come to be recognized as a standard tool of assessment of medical competencies. OSCE has been used for both formative and summative examination at graduate and postgraduate level, across the globe.

However, there is a Mr Hyde side to this Dr Jekyll. ***Table* II** outlines the factors that can affect the generalisabilty, validity, reliability and practicality of OSCE. The OSCE remains a toothless exercise if these factors are not taken care of. Unfortunately that is what is happening at most of the places where OSCE is now being introduced.

### Feasibility and Practicality

It is agreed that setting and running an OSCE is very resource intensive in terms of manpower, labor, time, and money; requires very careful organization; and meticulous planning(4). Training of examiners and patients, and preparation of stations and their checklists is a time consuming affair. Cost is high, both in human resource needs and money expended - patient (actor) payment, trainer payment, building

**TABLE I:** LIST OF MATERIALS NEEDED FOR THE CONDUCT OF OSCE

**General**

*VENUE:* Suitable spacious hall with sound proof partitions, or multiple adjacent rooms, waiting rooms for back up patients, rest rooms, refreshment area, briefing room

*FURNITURE:* Tables, chairs (for patient, examiner and examinee at each station), beds or examination couches, patient screen, signages, room heater or cooler

*TIMING DEVICE:* Stop watch or bell

*STATIONERY:* Score sheets, checklists, answer scripts, pens/ pencils

*MANPOWER:* Nurses, orderlies, simulated/ real patients, helpers/marshals

*CATERING:* Drinking water and food (snacks and lunch)

**Station Specific**

| Station No. | Station description | Basic equipment | Specific needs | Patient requirement |
|---|---|---|---|---|
| 1 | Data interpretation | Table, 1 chair | Calculator | – |
| 2 | Clinical examination of CNS | Patient screen, examination couch/ warmer, 2 chairs, heater/blower, handrub, paper napkins | Patellar hammer, cotton wisps, tuning fork, | 4 simulated patients, |
| 3 | Equipment: Phototherapy | Writing desk, 1 chair | Phototherapy equipment with duly labeled parts/components | – |
| 4 | Rest station | Table, 1chair | A tray with biscuits, napkins | – |
| 5 | Clinical photographs | Mounting board, writing desk, 1 chair | A chart with affixed and labeled photographs | – |

**TABLE II:** FACTORS AFFECTING THE USEFULNESS OF OSCE AS AN ASSESSMENT TOOL

| Factor | Limitation |
|---|---|
| Number of stations | Requires min 14-18 stations(1). Lesser the number-lesser the reliability(29), and lesser the content validity |
| Time for assessment | Lesser the time-lesser the reliability. A 10 minute station is more reliable as compared to a 5 minute station(25, 30) |
| Unreliably standardised patients | Limits reliability and validity |
| Individualised way of scoring | Limits reliability |
| Assessing only one component at a time | Limits validity(4) |
| Lack of item-analysis | Affects reliability(26) |
| Skill of the person preparing the checklist | May hamper objectivity; limits validity and reliability |
| Number of procedure stations | Lesser the number, lesser the clinical competencies that can be tested. Content specificity of stations limits reliability |
| Identification and deletion of problem stations | Increases reliability(26) |
| Task specific checklists | May not exactly replicate an actual clinical encounter, limits validity(13) |
| Blueprinting | Increases the content validity(10) |
| Competencies assessed | Not useful for assessing the learning behaviour, dedication to patients, and longitudinal care of patients(4) |
| Expensive and labor-intensive | Limits practicality and feasibility(31) |

rental or utilities, personnel payment, student time, case development, patient training, people to monitor, video taping etc. Most OSCEs are administered in medical center outpatient facilities. A separate room or cubical is needed for each station and this may be difficult to administer in smaller set-ups.

The problem is more acute in the developing countries and resource poor settings where a medical teacher has to assume the role of a consultant, service provider, researcher and administrator. This way, there is not much time the educator can spend on planning, preparing and executing an OSCE. This results in an OSCE which is more of an artefact and less of a true assessment.

**Objectivity**

The objectivity of OSCE is determined by the skill of the experts who prepare the OSCE stations and the checklists. Over the years, however, enthusiasm in developing detailed checklist (for increasing the objectivity) has led to another problem i.e. "trivialisation." The task is fragmented in to too many small components; and all of them may not be clinically relevant for managing a patient. A higher objectivity also does not imply higher reliability and that global ratings (which are by and large subjective) are a superior tool for assessment, especially in the hands of experienced examiners. An agreement has to be reached whether replacing the checklists by global rating on particular stations would improve the overall reliability, and then the OSCE can include both types of assessment tools.

**Validity**

Content validity can only be ensured by proper blueprinting(8). Following this, each task must be standardized and there must be itemization of its components using appropriate scoring checklists. Blueprinting also ensures multimodality OSCE that increases the content validity(19). Feedback from the examiners and the students can help in further improving the validity.

OSCE is not suited to assess the competencies related to characteristics like longitudinal care of patients, sincerity and dedication of the examinee to patient care and long-term learning habits (consequential validity)(20,21).

Mavis, *et al*.(21) have questioned the validity of OSCE by arguing that "observing a student perform a physical examination in OSCE is not performance based assessment unless data from this task is used to generate a master problem list or management strategy." Brown, *et al*.(23) have questioned the predictive and concurrent validity of OSCE by observing that the correlation between the students' result on OSCE and other assessment tools is low.

It would be appropriate to use OSCE to assess specific clinical skills (psychomotor domain) and combine it with other methods to judge the overall competency. Verma and Singh(24) concluded that OSCE needs to be combined with clinical case presentation for a comprehensive assessment. Panzarella and Manyon(25) have recently suggested a model for integrated assessment of clinical competence studded with supportive features of OSCE (ISPE: integrated standardized patient examination) to increase the overall validity.

**Reliability of OSCE on its own is less than desirable**

There are some issues related to reliability which need to be cleared for a proper understanding. Reliability does not simply mean reproducibility of results (for which, objectivity is a better term)- rather, reliability refers to the degree of confidence that we can place in our results (i.e. if we are certifying a student as competent, then how confident we are that he is really competent). This way of looking at reliability of educational assessment is different from the way we look at the reliability of say a biochemical test. It also needs to be understood that reliability is not the intrinsic quality of a tool; rather it refers to the inferences we draw from the use of that tool.

Reliability is generally content specific, meaning thereby that it is difficult to predict that if a student has done well on a case of CNS, he will do well on a case of anemia also.

Various factors can make results of OSCE less reliable include fewer stations, poor sampling, trivialization of the tasks, inappropriate checklists, time constraints, lack of standardized patients, trainer inconsistency, and student fatigue due to lengthy OSCEs. Leakage of checklists and lack of integrity of

both examiners as well as students can seriously compromise the validity as well reliability. A lot of variation has been reported when different raters have observed a station, and also between the performance from one station to another.

High levels of reliability (minimum acceptable defined as the reliability co-efficient of 0.8, maximum achievable: 1.0) can be achieved only with a longer OSCE session (of 4-8 h)(19). The reliability of a 1 and 2 h session is as low as 0.54 and 0.69, respectively; which is lower than the reliability of a case presentation of similar duration; but which can be increased to 0.82 and 0.9 in a 4 or 8 h session, respectively(21). However, it is impractical to conduct an OSCE of more than 3 hours duration. Newble and Swanson(26) were able to increase the reliability of a 90 min OSCE from 0.6 to 0.8 by combining it with a 90 minute free-response item written test.

Item analysis of OSCE stations and exclusion of problem stations is a useful exercise to improve the reliability(27). By ensuring content validity and by increasing the number of stations so that enough items can be sampled, reliability can be improved. All students should encounter similar test situation and similar real or simulated patients. Where it is difficult to arrange for similar real patients, it would be better to use simulated patients. However, arranging for children as simulated patients is usually not possible.

**Students' perception of OSCE**

Care should be exercised when introducing OSCE, specially if students have not experienced earlier (*e.g.* in basic sciences) because performing a procedure in front of an observer can be threatening to many students. Although the examiner is not required to say anything while observing the student, his/her body language can convey lot of anxiety. There have been reports to suggest that students do feel anxious initially(24,28). However, once explained the purpose and utility of direct observation in providing a good feedback and making learning better, acceptance is generally good.

**Traditional OSCE does not integrate competencies**

The OSCE model suggested by Harden revolves around the basic principle of "one competency-one task-one station." Skills were assessed in an isolated manner within a short time span. This does not happen in a real life scenario where the student has to perform all his skills in an integrated manner with the ultimate aim to benefit the individual and the community. The modern educational theory also stipulates that integration of tasks facilitates learning(21). It is thus imperative that the OSCE moves towards integrated assessment. For example dietary history taking and nutritional counseling can be integrated at one Station; similarly, chest examination and advising chest physiotherapy (based on the physical findings) can be integrated.

There are important implications of these aspects in the design of OSCE. There is a general agreement now that everything that is objective is not necessarily reliable; and conversely, all that is subjective is not always unreliable. It is also accepted that the advantages of OSCE do not relate to its objectivity or structure. If it was so, then the reliability of even a one hour OSCE would also have been high. Rather, the benefits seem to accrue from a wider sampling and use of multiple examiners, both of which help to overcome the threats to validity and reliability of assessment.

OSCE should not be seen as a replacement for something - for example, a case presentation or viva; rather it should be supplementing other tools. Using multiple tools helps to improve the reliability of assessment by taking care of content specificity and inter-rater variability. At the same time, one should not be over-enthusiastic to use OSCE type examination for competencies, which can be effectively tested by means of a written examination.

**INDIAN EXPERIENCES WITH OSCE**

OSCE has been by and large used as an assessment tool for formative assessment of undergraduate medical students at a few centers(5,24,29). Most of the faculty is not oriented to its use, and not many universities have incorporated it in summative assessment plan for the undergraduates. Probably this is because the Medical Council of India has yet to recognise and recommend it as an acceptable tool for summative assessment. Another main reason for

hesitancy, we feel, is the lack of training and time required on part of the faculty to initiate and sustain a quality OSCE.

National Board of Examination, Ministry of Health and Family Welfare, India has been using OSCE for summative assessment of postgraduate students for certification in the subjects of Otolaryngology, Ophthalmology, and Pediatrics for last few years. However, we feel that there are concerns as to the validity, reliability, scoring pattern and setting the standard in these examinations. For examples, there are only 6 procedure (observed) stations in a 24-30 station OSCE. The rest are based on recall and application of knowledge; for which more cost-effective testing tools are available. Many OSCE stations sample a very basic skill without relating them to a real life clinical situation. Most of the time, normal individuals are used as patient material. The standardized simulated patients include student nurse or a resident, who has not been trained specifically for this task. He/she is picked up only a few minutes before the exam. It is difficult to obtain uniformity in marking and inter-rater variability is likely to be more since the test is run concurrently at more than one center, spread all over India. There is no formal feedback given to the students or to the examiners to improve their performance. Finally, the passing standard is set arbitrarily at 50% which is not only not in conformity either with the accepted Angoff or Borderline approach but also obtained by adding the scores of multiple tools of variable reliability. Thus the OSCE pattern has limited validity and reliability and there is need for a re-look – either the present system be strengthened, or alternative methods should replace them.

## CONCLUSIONS

It is generally agreed that OSCE is a tool of assessment that tests competency in fragments and is not entirely replicable in real life scenarios. OSCE is useful for formative assessment, however on its own, it cannot be relied upon to fulfil the three necessary pre-requisites for a summative assessment as laid down by Epstein(30) i.e., promote future learning, protect the public by identifying incompetent physicians, and choosing candidates for further training. Limited generalizability, weak linkages to curriculum, and little opportunity provided for improvement in examinees' skill have been cited as the reasons for replacing OSCE with alternative methods in certain medical schools(22).

On a closer look there are gaps with respect to objectivity, validity and reliability of this assessment, especially in resource poor settings. It is costly and time consuming. It requires special effort and money to design OSCE stations needed to measure the essential professional competencies including ability to work in a team, professional ethical behavior, and ability to reflect on own (self-appraisal). For a summative assessment, OSCE should not constitute more than one-third of the total evaluation scheme and as far as possible, its grades should be reported separately. The need of the hour is an integrated multiplanar (3 dimensional) 360° assessment in its true perspective, of which OSCE can be a vital component.

*Funding*: None.

*Competing interests:* None stated.

## REFERENCES

1. Accreditation Council for Graduate Medical Education (ACGME). Outcome Project. http://acgme.org/Outcome/. Accessed on 23 July, 2009.
2. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002; 287: 226-235.
3. Harden RM, Stevenson W, Downie WW, Wilson GM. Assessment of clinical competence using an objective structured clinical examination. Br Med J 1975; 1: 447-451.
4. Harden RM, Gleeson FA. Assessment of clinical competence using objective structure clinical examination (OSCE). Med Edu 1979; 13: 41-54.
5. Gupta P, Bisht HJ. A practical approach to running an objective structured clinical examination in neonatology for the formative assessment of undergraduate students. Indian Pediatr 2001; 38: 500-513.
6. Boursicot K, Roberts T. How to set up an OSCE. Clin Teach 2005; 2: 16-20.
7. Waas V, van der Vleuten CPM. The long case. Med Edu 2004; 38: 1176-1180.

8. Newble D. Techniques of measuring clinical competence: objective structured clinical examination. Med Edu 2004; 38: 199-203.
9. Angoff WH. Scales, norms and equivalent scores. *In:* Thorndike RL, ed. Educational Measurement. Washington, DC: American Council on Education; 1971. p. 508-600.
10. Boursicot KAM, Roberts TE, Pell G. Using borderline methods to compare passing standards for OSCEs at graduation across three medical schools. Med Edu 2007; 41: 1024-1031.
11. Kaufman DM, Mann KV, Muijtjens AMM, van der Vleuten CPM. A comparison of standard setting procedures for an OSCE in undergraduate medical education. Acad Med 2001; 75: 267-271.
12. Dauphinee WD, Blackmore DE, Smee SM, Rothman AI, Reznick RK. Using the judgements of physician examiners in setting the standards for a national multi-center high stakes OSCE. Adv Health Sci Educ: Theory Pract 1997; 2: 201-211.
13. Smee SM, Blackmore DE. Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. Med Edu 2001; 35: 1009-1010.
14. Wilkinson TJ, Newble DI, Frampton CM. Standard setting in an objective structured clinical examination: use of global ratings of borderline performance to determine the passing score. Med Edu 2001; 35: 1043-1049.
15. Cusimano MD. Standard setting in medical education. Acad Med 1996; 71(suppl 10): S112-120.
16. Verma M, Singh T. Experiences with objective structured clinical examination (OSCE) as a tool for formative assessment in Pediatrics. Indian Pediatr 1993; 30: 699-702.
17. Reznick RK, Regehr G, Yee G, Rothman A, Blackmore D, Dauphinee D. Process-rating forms versus task-specific checklists in an OSCE for medical licensure. Acad Med 1998; 73: S97-99.
18. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998; 73: 993-997.
19. Walters K, Osborn D, Raven P. The development, validity and reliability of a multimodality objective structured clinical examination in psychiatry. Med Edu 2005; 39: 292-298.
20. Barman A. Critiques on the Objective Structured Clinical Examination. Ann Acad Med Singapore 2005; 34: 478-482
21. van der Vleuten CPM, Schuwirth WT. Assessing professional competence: from methods to programmes. Med Edu 2005; 39: 309-317.
22. Mavis BE, Henry RC, Ogle KS, Hoppe RB. The Emperor's new clothes: OSCE reassessed. Acad Med 1996; 71: 447-453.
23. Brown B, Roberts J, Rankin J, Stevens B, Tompkins C, Patton D. Further developments in assessing clinical competence. *In:* Hart IR, Harden RM, Walton HJ, eds. Further developments in assessing clinical competence. Montreal: Canadian Health Publications; 1987 .p. 563-571.
24. Verma M, Singh T. Attitudes of Medical students towards objective structured clinical examination (OSCE) in pediatrics. Indian Pediatr 1993; 30: 1259-1261.
25. Panzarella KJ, Manyon AT. A model for integrated assessment of clinical competence. J Allied Health 2007; 36: 157-164.
26. Newble D, Swanson D. Psychometric characteristics of the objective structured clinical examination. Med Edu 1988; 22: 325-334.
27. Auewarakul C, Downing SM, Praditsuwan R, Jaturatamrong U. Item analysis to improve reliability for an internal medicine undergraduate OSCE. Adv Health Sci Edu 2005; 10: 105-113.
28. Natu MV, Singh T. Student's opinion about OSPE in pharmacology. Indian J Pharmacol 1994; 26: 188-89.
29. Mathews L, Menon J, Mani NS. Micro-OSCE for assessment of undergraduates. Indian Pediatr 2004; 41: 159-163.
30. Epstein RM. Assessment in medical education. N Engl J Med 2007; 356: 387-396.