

Critiques on the Objective Structured Clinical Examination

A Barman,¹ MBBS (Bangladesh), DHE (India), MMed (UK)

Abstract

Introduction: The main aim of medical education is to foster the development of clinical competence in students at all levels. Differences in experiences, methods of instruction and ambiguous forms of assessment are obstacles to attaining this goal. Dissatisfaction with the conventional methods of clinical assessment on the part of teachers and students led assessors to search for appropriate alternatives and in 1975, Harden and his colleagues introduced the objective structured clinical examination (OSCE). It is nearly impossible to have a test that satisfies all the criteria of a good test. Sometimes, a compromise has to be made between the available resources (in terms of man, money and time), and the method and quality of assessment (in terms of reliability, validity, objectivity and practicability). **Methods:** This critique on the OSCE is based on the published findings of researchers from its inception in 1975 to 2004. **Results:** The reliability, validity, objectivity and practicability or feasibility of this examination are based on the number of stations, construction of stations, method of scoring (checklists and/or global scoring) and number of students assessed. For a comprehensive assessment of clinical competence, other methods should be used in conjunction with the OSCE. **Conclusion:** The OSCE can be a reasonably reliable, valid and objective method of assessment, but its main drawback is that it is resource-intensive.

Ann Acad Med Singapore 2005;34:478-82

Key words: Objectivity, Practicability, Reliability, Validity

Introduction

Assessment techniques appear to have an impact on study strategies¹ and to influence the performance of students.^{2,3} The proper selection of a method of assessment can improve student performance.² Faulty methods of assessment can lead to wrong decisions (pass or fail in certain areas) that might be deleterious to the future activity of students and the welfare of the community.⁴

Both teachers and students were dissatisfied with the conventional methods of clinical assessment.⁵ This dissatisfaction led assessors to search for appropriate alternatives.⁶ In 1975, Harden and his colleagues introduced the objective structured clinical examination (OSCE), claiming that it fulfilled all the criteria of an ideal method of assessment of clinical competence.⁷

Four criteria, comprising validity, reliability, objectivity

and practicability or feasibility, must be considered in developing any method of assessment. This critique on the OSCE is based on the published findings of a large number of researchers from its inception in 1975 to 2004.

Methods

A search of PubMed, with the key words “validity, reliability of OSCE”, was conducted. The latest article found at the beginning of writing the manuscript was from Newble;⁷ all other articles and books referred to in this article are subsequently cross-searched back to 1975, when a preliminary report describing the OSCE was published in the British Medical Journal.⁸

Results

This critique is a description of the features of the OSCE in terms of its reliability, validity, objectivity and feasibility.

¹ Department of Medical Education
School of Medical Sciences
Universiti Sains Malaysia, Malaysia

Address for Reprints: A/Prof Arunodaya Barman, Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia.

Email: barman@kb.usm.my

Validity

Newble⁴ stated that validity is the most important criteria of an assessment method. The validity of a test refers to the extent to which it measures what it is intended to measure. To ensure that a test has a high level of validity, it must contain a representative sample of what the students are expected to have achieved both in terms of subject matter and educational objectives. Face validity or credibility refers to the extent to which the OSCE includes material that is considered important to measure. A valid clinical examination should assess the components of clinical competence from history-taking to management and patient education.⁹

Content validity refers to the degree to which the OSCE covers the area of competence in question. The OSCE fulfils all the criteria of a good test, including that of validity.⁹ To make the OSCE maximally valid, a test matrix was necessary. Newble et al¹⁰ suggested a test blueprint to achieve such content validity.

Brown et al¹¹ stated that the predictive and concurrent validity of the OSCE are low. Concurrent validity compares the evaluation of performance of a task by the OSCE to the evaluation of the same task by the best existing external measure available. When they explored the concurrent validity, the correlations between the students' results on the OSCE and their results on other courses or grades were low. Cox¹² stated that there is no evidence that performance-based tests such as the OSCE are of greater validity than traditional written or oral examinations. Norman¹³ argued that cognitive measures of clinical competence are more psychometrically efficient than performance-based measures. Competencies relating to characteristics such as the longitudinal care of patients, dedication to patients, and long-term ability to continue learning cannot be efficiently assessed by the OSCE and have to be assessed by other methods. Information provided by an OSCE for assessment decision is complementary to that obtained from other sources.¹⁴ Verma and Singh¹⁵ felt that the OSCE could assess specific clinical skills but other methods should be used in conjunction with the OSCE for a comprehensive assessment of clinical competence. In the OSCE, the ability to take a full, detailed history for a particular case cannot be tested because of the time limitation. The skills and knowledge of a student are assessed in parts, and students cannot be tested on their ability to look at a patient as a whole.⁹

The OSCE is limited in its ability to measure what the student would do in real-life situations for the care of a patient.¹⁶ The OSCE purports to show the students' skills by examining a number of patients in isolation instead of comprehensively examining a single patient.¹⁷ Some of the respondents in Newble's study found that the OSCE

examines a narrow range of knowledge and skills and does not test for history-taking competence properly.¹⁸ Clinical competence includes a number of attributes. The OSCE alone cannot assess all these attributes efficiently,¹⁹ unless it is combined with some other assessment methods. Measuring all the parts in isolation is not equivalent to measuring the whole integrated performance. Diagnostic procedures are not the sum of all the partial skills learned as routines and tested in structured examinations.¹² Moreover, diagnostic exploration is generally selective. Clinical diagnosis-searching moves selectively. Busy clinicians follow a principle of least effort, aiming for efficiency by carrying out the minimum steps needed to reach a diagnosis, rather than completing whole activities in series and sequence. Assessing students on a fixed set of components as demanded by the OSCE seems thorough, but it penalises the shrewdness and efficiency of those who will be able to diagnose with the least effort. Cox¹² also stated that the OSCE demands that students conform to the structured path set by the test's creators. In effect, the focus of the examination becomes its structure, rather than the learners.

Patil²⁰ stated that although the OSCE seems to assess knowledge and skills of undergraduate medical students, it does not assess the in-depth knowledge and skills that are necessary for postgraduate students.

Reliability

Reliability refers to the reproducibility of a set of measurements, consistency or stability of measures over time and over test forms, including different samples of items. Time, especially short time periods that require speedy response, may greatly affect reliability. The reliability of OSCE is low if a few stations²¹ and little time are used in assessment.¹⁸ Other factors of low reliability are unreliably standardised patients, personalised ways of scoring, and disorganised staff or noisy rooms.²² For a high level of reliability, OSCEs must have a large number of stations, and be combined with other methods of assessment.⁷ Brown et al¹¹ found that the internal consistency of a total OSCE as measured by Cronbach alpha was 0.50. Unreliability in the clinical examination may arise from the fact that different students are assessed on different patients and one may come across a temperamental patient who may help some students while obstructing others. This may happen in the case of OSCE when a multiple circuit session needs to be organised to assess a large number of students at a time. Harden and Gleeson⁹ suggested an ideal solution for that: all students would examine all the patients in the OSCE.

The scores of a test may not accurately reflect students' ability as repetitious demands may fatigue the student,

patient or examiner. Students' fatigue due to lengthy OSCEs may affect their performance.^{23,24} A heavy workload may affect examiners' performance.²⁵ Moreover, some students experience greater tension before and during examinations, as compared to other assessment methods.²⁶ In spite of efforts to control patient and examiner variability, inaccuracies in judgment due to these effects remain.²⁷ Gledhill and Capotos²⁷ observed that there was no correlation in the performances of students in two examinations; many students who did well in one examination failed in another. They concluded that this happened partly due to patient and examiner variability, and the test sequence. Even after taking care of all sources of measurement error related to candidates, cases, raters and standardised patients, reliability coefficients of OSCEs have been reported to be between 0.41 and 0.88.^{23,28} The accuracy of standardised patient simulation²⁹ and the order in which stations are encountered³⁰ also affect students' performance. Newble et al³¹ stated that oral examinations, when properly conducted, can be as reliable as the OSCE.

Objectivity

The OSCE claims to be objective. Objectivity rests on the standardisation of the task and the scoring checklist for the stations.⁸ There is evidence that this is not always the case, especially for some types of stations. The global rating scale scores, given by expert examiners, showed higher inter-station reliability and better predictive validity than did the scores using checklist.³² It has been shown in a number of studies that differences in station-to-station performance are much greater than those between raters observing the same station.³³⁻³⁵ Newble and Swanson³⁶ calculated interrater reliabilities in their study as Pearson correlations between examiner pairs and they found that these varied, both within and between stations. Pearson correlation for 10-minute physical examination stations were - 0.72, 0.72, 0.75, 0.82 and 0.89; for 5-minute physical examination stations - 0.33, 0.45, 0.64 and 0.76; patient education stations - 0.38, 0.39, 0.55 and 0.75; procedural skills stations - 0.48, 0.52, 0.76 and 0.91. Inter-station correlation was computed using Spearman-Brown prophecy formula and these values for the patient stations and static stations were 0.08 and 0.06, respectively. These coefficients are much lower than the interrater reliabilities. This indicates that good interrater reliability requires a larger number of stations. Though the OSCE was developed to decrease bias in the assessment of clinical competence, it is not without the pitfalls of other types of examinations.³⁷

Practicability/Feasibility

In addition to the aforementioned factors, the practicality, or the feasibility of the test method needs to be taken into consideration as well. In selecting an assessment method

one should consider the number of students to be assessed; the number of staff, status and their specialties; the availability of patients; space;⁸ and time and money.

The OSCE, compared to the structured oral (SO) examination and other traditional examinations, is more time-consuming and more expensive in terms of human and material cost.^{38,39} The time required for setting up the examination is greater than that needed for the traditional assessment methods.^{8,40}

To assess 117 students in a 22-station OSCE with 93 simulated patients required 212 man-hours, hence the total time spent for the examination was 527 man-hours.⁴¹ The OSCE is not a totally reliable method^{34,36} if it is not approximately a 6-hour session,²³ which can be impractical, especially when real patients are used.

The OSCE costs \$15 per student for each hour of the examination, which is more than the cost of a test conducted by the National Board of Medical Examination (\$10 per student per hour).² Stillman et al³³ estimated that a per student testing cost of about \$200 would be needed to achieve an acceptable level of reliability. Grand'Maison et al⁴² revealed that it costs \$1080 to assess the performance of each student in the Quebec Medical College. Expenses for simulated patients were also high, costing approximately \$7 per simulated patient per student test.⁴³ In 1991, Ainsworth et al¹⁶ reported that the University of Texas Medical Branch spent \$10 per hour per standardised patient training or participation in examination but simulated patients undergoing rectal examination were paid \$20 to \$30 per hour.

The OSCE is costly and labour-intensive.⁴⁴⁻⁴⁷ These are the drawbacks which make it difficult for many schools to use the OSCE to assess their students.⁴⁴

Blackwell et al⁴⁸ identified that it is difficult to administer the OSCE, and that scoring requires a lengthy period of time. Richards et al⁴⁹ stated that manual scoring of OSCE stations is time-consuming and increases the probability of mistakes. They suggested computer scoring, where coded answers are read by optical scanners.

For maintaining the reliability of the OSCE, all students must be exposed to similar test situations. In history-taking and physical examination stations, the same or similar real patients or simulated patients are required.³¹ It is very difficult to have a number of similar real patients. To overcome this problem, simulated standardised patients can be used, but training them for use in test situations incurs costs and manpower.¹⁶ It is nearly impossible to have children as standardised patients or patients with similar physical findings.⁴²

A single OSCE for all students would be the best, but it would stretch resources in terms of examiners, patients and

accommodation. It may be very difficult to use the OSCE to assess a large number of students with a number of cases at multiple centres.⁴² Mujir⁵⁰ reported that Unversiti Sains Malaysia solved its problems in implementing a large-scale OSCE to assess a large number of its students by triplicating the OSCE circuits, accommodating the stations in multidisciplinary laboratories and inviting clinicians from the local general hospital to be examiners. Successful administration of OSCE very much depends on the harmonious activity of all concerned, since the absence of even a single patient or examiner will stop or delay the process.⁵¹

Conclusion

The OSCE is developed to reduce bias in the assessment of clinical competence, but it is not without the pitfalls of other assessment methods. It is nearly impossible to have an assessment method that satisfies all the criteria of a good test. Despite some reservations relating to reliability, validity, objectivity and practicability, the OSCE is now an established part of the repertoire of clinical assessment skills in many medical schools around the world. To ensure a reasonable level of validity, reliability and objectivity, great care needs to be taken in its planning and administration. By using of a test matrix, spreading the learning objectives that need to assessed (identified in the test matrix) over a number of stations, accommodating the stations in available places, training the academic staff and using appropriate checklist, the OSCE can be made into a reliable, valid, objective and practicable test method for any medical institution.

REFERENCES

1. Marton F, Saljo R. On qualitative differences in learning: II—outcomes as a function of the learner's conception of the task. *Br J Educ Psychol* 1976;46:115-27.
2. Petrusa ER, Blackwell TA, Rogers LP, Saydjari C, Parcel S, Guckian JC. An objective measure of clinical performance. *Am J Med* 1987;83:34-42.
3. Newble DI, Entwistle NJ. Learning styles and approaches: implications for medical education. *Med Educ* 1986;20:162-75.
4. Newble DI. Assessing clinical competence at the undergraduate level. *Med Educ* 1992;26:504-11.
5. Simpson MA. *Medical education: A critical approach*. London: Butterworths, 1972:118-57.
6. Newble DI, Baxter A, Elmslie RG. A comparison of multiple-choice tests and free-response test in examinations of clinical competence. *Med Educ* 1979;13:263-8.
7. Newble DI. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ* 2004;38:199-203.
8. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examinations. *BMJ* 1975;1:447-51.
9. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13:41-54.
10. Newble DI, Hoare J, Elmslie RG. Validity and reliability of a new examination of the clinical competence of medical students. *Med Educ* 1981;15:46-52.
11. 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, editors. *Further developments in assessing clinical competence*. Montreal: Canadian Health Publications, 1987:563-71.
12. Cox K. No Oscar for OSCA. *Med Educ* 1990;24:540-5.
13. Norman GR. Reliability and construct validity of some cognitive measures of clinical reasoning. *Teaching Learning Med* 1989;1:194-9.
14. Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. *Arch Intern Med* 1990;150:573-7.
15. Verma M, Singh T. Experiences with objective structured clinical examination (OSCE) as a tool for formative evaluation in pediatrics. *Indian Pediatr* 1993;30:699-702.
16. Williams RG, Barrows HS. Performance based assessment of clinical competence using clinical encounter multiple stations. In: Hart IR, Harden RM, Walton HJ, editors. *Further developments in assessing clinical competence*. Montreal: Canadian Health Publications, 1987:125-33.
17. Ainsworth MA, Rogers LP, Markus JF, Dorsey NK, Blackwell TA, Petrusa ER. Standardized patient encounters. A method for teaching and evaluation. *JAMA* 1991;266:1390-6.
18. Newble DI. Eight years' experience with a structured clinical examination. *Med Educ* 1988;22:200-4.
19. Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. *Med Educ* 1991;25:293-9.
20. Patil JJ. Objective structured clinical examination. *CMAJ* 1993;149:1376-8.
21. Wessel J, Williams R, Finch E, Gémus M. Reliability and validity of an objective structured clinical examination for physical therapy students. *J Allied Health* 2003;32:266-9.
22. Smee S. Skill based assessment. *BMJ* 2003;326:703-6.
23. Van der Vleuten C, Swanson DB. Assessment of clinical skills with standardized patients: State of the art. *Teaching Learning Med* 1990;2:58-76.
24. Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ. Student fatigue as a variable affecting performance in an objective structured clinical examination. *Acad Med* 1990;65(9 Suppl):S53-4.
25. Reznick R, Smee S, Rothman A, Chalmers A, Swanson D, Dufresne L, et al. An objective structured clinical examination for the licentiate: report of the pilot project of the Medical Council of Canada. *Acad Med* 1992;67:487-94.
26. Kirby RL, Curry L. Introduction of an objective structured clinical examination (OSCE) to an undergraduate clinical skills programme. *Med Educ* 1982;16:362-4.
27. Gledhill RF, Capatos D. Factors affecting the reliability of an objective structured clinical examination (OSCE) test in neurology. *S Afr Med J* 1985;67:463-7.
28. Swanson DB, Norcini JJ. Factors influencing reliability of tests using standardized patients. *Teaching Learning Med* 1989;1:158-66.
29. Tamblyn RM, Klass DK, Schanbl GK, Kopelow ML. Factors associated with the accuracy of standardized patient presentation. *Acad Med* 1990;65(9 Suppl):S55-6.
30. Lloyd JS, Williams RG, Simonton DK, Sherman D. Order effects in standardized patient examinations. *Acad Med* 1990;65:S51-2.

31. Newble DI, Hoare J, Sheldrake PF. The selection and training of examiners for clinical examinations. *Med Educ* 1980;14:345-9.
 32. 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;73:993-7.
 33. Stillman PL, Swanson DB, Smee S, Stillman AE, Ebert TH, Emmel VS, et al. Assessing clinical skills of residents with standardized patients. *Ann Intern Med* 1986;105:762-71.
 34. Swanson DB. A measurement framework for performance based tests. In: Hart IR, Harden RM, Walton HJ, editors. *Further developments in assessing clinical competence*. Montreal: Canadian Health Publications, 1987:13-45.
 35. Norman GR, Muzzin LJ, Williams RG, Swanson DB. Simulation in health sciences education. *Journal of Instructional Development* 1985;8:11-7.
 36. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. *Med Educ* 1988;22:325-34.
 37. Soeters D, Scherpbier AJJA, van Lunsen HW. Assessing students' performances or examination peculiarities. In: Hart IR, Harden RM, Walton HJ, editors. *Further developments in assessing clinical competence*. Montreal: Canadian Health Publications, 1987:517-23.
 38. Cusimano MD, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R. A comparative analysis of the costs of administration of an OSCE (objective structured clinical examination). *Acad Med* 1994;69:571-6.
 39. Cohen R, Reznick RK, Taylor BR, Provan J, Rothman A. Reliability and validity of the objective structured clinical examination in assessing surgical residents. *Am J Surg* 1990;160:302-5.
 40. Smith LJ, Price DA, Houston IB. Objective structured clinical examination compared with other forms of student assessment. *Arch Dis Child* 1984;59:1173-6.
 41. Frye AW, Richards BF, Philp EB, Philp JR. Is it worth it? A look at the costs and benefits of an OSCE for second-year medical students. *Med Teach* 1989;11:291-3.
 42. Grand'Maison P, Lescop J, Rainsberry P, Brailovsky CA. Large-scale use of an objective, structured clinical examination for licensing family physicians. *CMAJ* 1992;146:1735-40.
 43. Williams RG, Barrows HS, Vu NV, Verhulst SJ, Colliver JA, Marcy M, et al. Direct standardized assessment of clinical competence. *Med Educ* 1987;21:482-9.
 44. Reznick RK, Smee S, Baumber JS, Cohen R, Rothman A, Blackmore D, et al. Guidelines for estimating the real cost of an objective structured clinical examination. *Acad Med* 1993;68:513-7.
 45. Kowlowitz V, Hoole AJ, Sloane PD. Implementing the objective structured clinical examination in a traditional medical school. *Acad Med* 1991;66:345-7.
 46. Donnelly MB, Sloan DA, Rubeck RF, Schowartz RW. Performance-based testing in medical education. *Curr Surg* 1992;49:566-72.
 47. Sloan DA, Donnelly MB, Schwartz RW. Assessing the problem-based clerkship: the rationale for and cost of a multidimensional assessment program. *Med Teach* 1994;16:11-8.
 48. Blackwell TA, Ainsworth MA, Dorsey NK, Callaway MR, Rogers LP, Collins KE. A comparison of short-answer and extended matching question scores in an objective structured clinical exam. *Acad Med* 1991;66(9 suppl):S40-2.
 49. Richards BF, Philp EB, Philp JR. Scoring the objective structured clinical examination using microcomputer. *Med Educ* 1989;23:376-80.
 50. Mujir MI. Problems and their solutions to effective large scale implementation of the Objective Structured Clinical Examination (OSCE); the Universiti Sains Malaysia experience. In: Hart IR, Harden RM, Walton HJ, editors. *Further developments in assessing clinical competence*. Montreal: Canadian Health Publications, 1987:637-9.
 51. Harden RM. Twelve tips for organizing an objective structured clinical examination (OSCE). *Med Teach* 1990;12:259-64.
-