

Barriers to Appropriate Standard Performance of an Objective Structured Clinical Examination

Since 1960s, many experts criticized the use of traditional examinations to assess the skills of medical trainees at all levels and many efforts implemented to create a feasible instrument to measure these skills. Objective Structured Clinical Examination (OSCE) is a modern type of evaluation which is used in many psychiatric and other clinical centers worldwide and has shown a credible validity and reliability. This editorial highlights controversies among advocates and opponents and barriers to appropriate standard performance of this method which has been used in the oral examination of the Board of Psychiatry in Iran since 2006.

In the 7th decade of 19th century, data gathered by the National Board of Medical Examination in the United States of America, involving over 10,000 medical students showed that the correlation of independent evaluations by two examiners was less than 0.25 (1). In the next years, a large body of literature has criticized the use of an oral examination and Individual Patient Assessment (IPA) to assess the skills of medical trainees at all levels and concluded that the traditional examinations have poor reliability and validity (2-7).

Many efforts implemented to create a valid and reliable instrument to measure clinical skills of medical students. A valid and reliable examination must be free of bias and distortion. Reliability and validity are two important concepts for defining and measuring bias and distortion. Reliability refers to the extent to which assessments are consistent. Instruments such as national standardized exams should not make any difference whether a trainee takes the assessment in the morning or afternoon; one day or the next; with one examiner or another one. Validity of an examination refers to whether or not it measures what it is supposed to measure. Validity and reliability of an important assessment, such as the national board exam, is essential (8).

Among the clinical skills necessary for the professional development of medical students- especially in psychiatry training- is the ability to create a doctor-patient relationship through effective communication skills (9), as well as diagnosis of symptoms of the disorders and therapeutic skills and procedures.

Although some of these clinical skills can be measured without a patient, most of them should be assessed in the presence of a patient. An ideal assessment setting is in the presence of real patients. But it is not possible and ethical to examine many patients with many examinees in the presence of same examiners.

Objective Structured Clinical Examination (OSCE) is a modern type of evaluation (10) often used in medical sciences (11) to measure clinical skills such as communication, clinical examination, medical procedures / prescriptions / techniques, and interpretation of results (12).

This method which is used in many psychiatric and other clinical centers worldwide (13-15) has shown a credible validity and reliability (16-18).

OSCE normally consists of several short stations, in which each examinee is examined on a one-to-one basis with one or several examiner(s) and either real (19) or simulated patients (12), computer simulators (20), moulages (21) or videos (22). It is considered to be an improvement over traditional assessment methods because the objective stations can be standardized enabling fairer peer comparisons and complex procedures can be assessed without endangering patients' health. There are criticisms that the OSCE stations can never be truly standardized and objective in the same way a written Multiple Choice Question (MCQ) can (12). However, some investigations indicate that MCQ marks correlate significantly with the OSCE marks (23).

The OSCE also incorporates the technology of standardized patients (SPs) (24). Standardized patients (also referred to as simulated patients, patient instructors, programmed patients, pseudo or surrogate patients) (25-29), are individuals, either with or without "real" symptoms, who are coached by medical educators to act as patients for teaching and/or assessment purposes (25,30).

The use of standardized patients allows the nature of problems and the level of difficulty to be standardized for all students (31), so it is one of the most important components of OSCE. If the station's approach is role playing of SPs, one should use a person who has been properly trained for this task (32). SPs are not a homogeneous group (33). There have been recruited people from "professionally trained actors" (28) to "drama students and housewives, some of whom had no previous acting experience" (34) or "the allied health programs at local colleges, community volunteer programs the community senior citizen program and from clinics at a university-based department of family medicine" (29). Their only common characteristic is that of simulating real patients. This raises the question of whether such a diverse group can be trained to behave in a 'standardized' way (33). Extensive research has been conducted to support these testing initiatives in identifying individuals with clinical skills deficiencies. They provide a framework for validating new simulation modalities and extending simulation-based assessment into various fields of medicine (35). It is worth mentioning that the often highly emotional nature of SPs' roles may have a residual effect on the simulators (16).

Regarding to its reliability, another important aspect of OSCE is preparation of a standard printed checklist or blueprint for each station that the examiners have to fill out while observing the examinee. These checklists are standardized to reduce examiners' bias. There are station specific points and general performance points to be evaluated by these checklists. Some experts believe that there are up to 40 points to be checked in each station (36). Most SP checklists are developed in one of three ways: by case writers, by experts reacting to checklists proposed by case writers, or by a panel of experts (37). Typically, a SP checklist or blueprint represents an agreed-on selection of critical items that an examinee should address in a specific encounter. However, for a particular encounter, even among

experts there may be extreme variance in the specific items addressed. So to say, the checklist items include only those items on which everyone agrees (38).

OSCE has specific costs associated with the use of SPs, as well as materials, supplies, and the time of examiners and support staff (39). Many efforts implemented to create organizations of healthcare providers, educators, engineers, and scientists to achieve their common goal of furthering healthcare education, patient safety and patient-centered care using simulation in all forms to advance the interests of professionals who make use of SPs in healthcare education and assessment, such as the Association for Standardized Patient Educators (ASPE) and the Society for Simulation in Healthcare (SSH) (40,41). Efforts have been done to improve standardization among SPs (42), and many workshops have been planned to practice interview, communication and performance skills (43-46), assistive/ useful feedback skills to medical students (47,48), distortion factors (49) and so on (50).

Although OSCE's reliability and validity is more difficult to achieve in psychiatry examinations, it has been used in the oral examination of the Board of Psychiatry in Iran since 2006 (32). But there are many controversies among advocates and opponents. Opponents indicate to their own experience and some previous studies and believe that the OSCE is not a feasible stand-alone method for summative evaluation of clinical practice. They believe that the expense of SPs, whose training and time spent performing accounted for the largest proportion of the direct cost of setting up OSCEs is a major disadvantage (16). Opponents also argue that SPs may become ideal "textbook cases" who play idealized "textbook scenarios", to which real patients, with all their idiosyncrasies, do not often conform. So, they may not mimic real-life situations and OSCE may not allow assessment of complex skills (33). They believe that the binary checklists usually used in most OSCEs are insufficiently sensitive to capture differences between novices, acceptable performers, and experts to detect higher clinical components such as rapport, empathy and ethics (38,51). Their next objection reason is that although a typical OSCE station lasts up to 10-15 minutes, a traditional psychiatric interview is about 50 minutes. This important issue raises questions about the content validity of the short OSCE station (33). Finally, indicating that some complex psychiatric presentations such as thought disorder are difficult to simulate (17), opponents suggest that this performance-based evaluation method should be combined with other more valid evaluation methods.

Advocates argue that combination of multiple observations and standardization of content and difficulty offers excellent psychometric properties for the OSCE. They believe that traditional measurements are technically challenging and carry ethical issues of patient confidentiality and vulnerability, and the luck of the draw in selection of examiner and patient play a significant role in the outcome of postgraduate oral traditional examinations in psychiatry (52). They also indicate that an advantage of SPs over real patients is that of allowing different students to be presented with a similar challenge (patient variable in examination is uniform across trainees), thereby reducing an important source of variability (53). Other advantages that the advocates describe include SPs' reliable availability and adaptability, which enables the reproduction of a wide range of clinical phenomena tailored

to the student's level of skill. In addition, SPs can simulate scenarios that may be distressing for a real patient, such as terminal illness or bereavement (54).

It seems that the major base of controversies is the lack of appropriate budget, recruitment of appropriate SPs and organizations for professional training of them that are barriers to appropriate standard performance of OSCE. Amateur SPs who have not been trained properly with professional trainers cannot do the simulation appropriately. A key requisite for achieving both accuracy and consistency of SP performance is good training (33). SPs need to draw on their own personal experiences with physicians, conversations with healthcare professionals, talking to specific real patient populations and so on. They also need to be trained to accurately and reliably simulate particular clinical scenarios to act with high validity. Frequent quality assessment is needed to ensure consistency in the portrayal of the patient role, especially since SPs may absorb quite a significant amount of clinical knowledge during their interactions with physicians, other healthcare professionals (55) and real patients as well as professional trainers. Multiple encounters may be needed for broad ranged training or testing. Standard methods have been described by Barrows (56). For example, the work of an experienced SP trainer has been described as: 'training for a role begins with the presentation of written material and where possible, video footage of real patients. Each SP is then observed performing the role by the station's author to verify the realism of the portrayal and to ensure consistency across the SPs in their presentation of affect and in their response to questions' (16). As you see, recruitment of SPs is difficult, time consuming and more expensive than using "real" patients (57). Cost of training and application and use of resources are major barriers to the extensive use of SPs worldwide. Appropriate quarantine is also needed for standardized patients in the case of assessment setting.

At the same time, SPs are case specific and are able to assess clinical competency in a limited area only. Also, while appropriately trained SPs are quite proficient in simulating the symptoms, mental states and even certain physical examination findings, they may not be able to simulate certain other signs such as fever, heart murmurs or lung sounds (57). This problem is less prominent in the field of psychiatry and may partially be resolved by computer simulators, videos or mouldages in other medical settings.

Content specificity is a fundamental problem in assessing clinical problem solving ability with SPs. This is a characteristic of problem solving in real clinical experience, so it can be expected on SPs as well. Some experts recommend abandoning checklists or at least rethinking our approach to creating checklists and supplementing checklists with other measures for better assessment of cognitive skills and problem solving abilities. Supplementation with global assessment by an examiner may improve testing characteristics (38,58). It is necessary to use large numbers of cases to adequately evaluate problem solving ability (59). An individual real patient assessment or evaluating clinical performance on a single SP with more accuracy and fidelity may well result in a less valid test because many technical skills would be failed to be evaluated.

Another barrier to Sp's participation in the OSCE is shame. But it can be mitigated (60). Finally, it is not uncommon at certain institutions for members of teaching staff known to students and vice versa to assess students. This need not affect the integrity of the examination process although there is a deviation from anonymous marking (12).

It is important to mention that OSCEs' nature is to break down clinical skills into small 'testable' tasks. Doctors who are very good at performing these piecemeal tasks without being able to assimilate them into a coherent assessment will pass the OSCE successfully. Wallace et al resemble them to a pianist who can play beautiful scales and arpeggios but cannot play a complete sonata (33). They hope that senior candidates preparing for OSCEs do not forget how to take a history, make a diagnosis and formulate a management plan. Even though this will be tested in Part II of the examinations, they feel these are essential skills for all doctors, from house officer to consultant (33). The reason of many false positive and false negative diagnoses in OSCE is the lack of a holistic approach.

In conclusion, research has shown that reliability and validity of OSCE can be achieved in psychiatric assessment, but it depends on adequate and standardized SP training methods, preparing checklists, and supplementing them with other measures such as supplementation with global assessment by an examiner. However, some additional evaluations may be needed.

References

1. Hubbard JP, Levit EJ, Schumacher CF, Schnabel TG. An objective evaluation of clinical competence: new techniques used by the National Board of Medical examiners. *N Engl J Med* 1965; 272: 1321-8.
2. Hodges B, Regehr G, Hanson M. An objective structured examination for examining clinical clerks. *Acad Med* 1997; 72(8):715-21.
3. Abrahamson S. The oral examination: the case for and against, in *Evaluating the Clinical Skills of Medical Specialists*. Evanston, IL, American Board of Medical Specialties; 1983. p. 121-4.
4. McGuire CH. The oral examination as a measure of professional competence. *J Med Educ* 1966; 41(3):267-74.
5. Davidson RG. A point of view: oral examinations (letter). *Ann R Coll Physicians Surg Can* 1983; 16:114.
6. Valberg LS, Stuart RK. A point of view: university in-training evaluation and oral examinations in internal medicine. *Annals RCPSC* 1983; 16:513-5.
7. Jayawickramarajah PT. Oral examinations in medical education. *Med Educ* 1985; 19(4): 290-3.
8. Classroom Assessment of clinical competence using objective structured examination: assessment, basic concepts. Available from: URL: <http://fcit.usf.edu/assessment/sitemap.html>

9. Bolton S, Harward D, Smith A, Camp G, Petrusa E, Richards B, et al. Relation skills in a clinical performance examination: reliability and validity of the relationship instrument. [Dissertation]. American Educational Research Association; 1995.
10. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J* 1975; 22; 1(5955): 447-51.
11. Ross M, Carroll G, Knight J, Chamberlain M, Fothergill-Bourbonnais F, Linton J. Using the OSCE to measure clinical skills performance in nursing. *J Adv Nurs* 1988; 13(1): 45-6.
12. The Wikipedia Organization. Wikipedia, the free encyclopedia 2008; Available from: URL: <http://en.wikipedia.org/wiki/>
13. Hodges B, Hanson M, Mc Naughton, Regehr G. Creating, monitoring and improving a psychiatry OSCE: a guide for faculty. *Acad Psychiatry* 2002; 26(3): 134-61.
14. Yudkowsky R. Should we use standardized patients instead of real patients for high stakes exams in psychiatry. *Acad Psychiatry* 2002; 26(3): 187-91.
15. Michael A. OSCE in Psychiatry. New York: Churchill Livingstone; 2004.
16. Hodges B, Regehr G, Hanson M. An objective structured clinical exam for evaluating Psychiatric clinical clerks. *Acad Med* 1997; 72(8): 715-21.
17. Hodges B, Regehr G, Hanson M. Validation of an objective structured clinical examination in psychiatry. *Acad Med* 1998; 73(8): 910-2.
18. Park RS, Chibnall JT, Blaskiewicz RJ, Furman GE. Construct validity of an objective structured clinical examination (OSCE) in Psychiatry: association with the clinical skills examination and other indicators. *Acad Psychiatry* 2004; 28(2): 122-8.
19. McClure CL, Gall E, Meredith K E, Gooden MA, Boyer JT. (1985) Assessing clinical judgement with standardized patients. *J Fam Pract* 1985; 20(5): 457-64.
20. Courteille O, Bergin R, Stockeld D, Ponzer S, Fors U. The use of a virtual patient case in an OSCE-based exam- a pilot study. *Med Teach* 2008; 30(3): e66-76.
21. Möhrle M, Jürgens S, Zipfel S, Schrauth M. Modern investigation with historical methods. Objective structured clinical examination (OSCE) of moulages. *Hautarzt* 2006; 57(6): 528-31.
22. Cartney P. Using Video Interviewing in the Assessment of Social Work Communication Skills. *Br J Soc Work* 2006; 36(5): 827-44.
23. Famuyiwa OO, Zachariah MP, Ilechukwu STC. The objective structured clinical examination in undergraduate psychiatry. *Med Educ* 1991; 25(1): 45-50.
24. Barrows HS, Abrahamson S: The programmed patient: a technique for appraising student performance in clinical neurology. *J Med Educ* 1964; 39: 802-5.
25. Chur-Hansen A, Koopowitz L. Introducing Psychosocial and Psychiatric Concepts to First Year Medical Students Using an Integrated, Biopsychosocial Framework. *Educ Health* 2002; 15(3): 305-14.
26. Badger LW, DeGruy F, Hartman J, Plant MA, Leeper J, Ficken R, et al. Stability of standardized patients' performance in a study of clinical decision making. *Fam Med* 1995; 27(2): 126-33.
27. Sanson-Fisher RW, Poole AD. Simulated patients and the assessment of medical students' interpersonal skills. *Med Educ* 1980; 14(4): 249-53.
28. Norman GR, Tugwell P, Feightner J W. A comparison of resident performance on real and simulated patients. *J Med Educ* 1982; 57(9): 708-15.

29. Rubin NJ, Philp EB. Health care perceptions of the standardized patient. *Med Educ* 1998; 32(5): 538-42.
30. Vu NV, Barrows HS. Use of standardized patients in clinical assessments: recent developments and measurement findings. *Educ Res* 1994; 23(3): 23-30.
31. Vander vleuten CPM, Swanson D. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990; 2: 58-76.
32. Taghva A, Bolhari J, Bahador H, Attari A, Fadai F, Khademi M, et al. How to run psychiatry OSCE in three days. *Iranian Journal of Psychiatry and Behavioral Sciences* 2008; 2(2): 5-9.
33. Wallace J, Rao R, Haslam R. Simulated patients and objective structured clinical examinations: review of their use in medical education. *Advances in Psychiatric Treatment* 2002; 8: 342-8.
34. Jason H, Kagan N, Werner A, Elstein AS, Thomas JB. New approaches to teaching basic interview skills to medical students. *Am J Psychiatry* 1971; 127: 140-2.
35. Boulet, JR, Smee, SM, Dillon GF, Gimpel JR. The Use of Standardized Patient Assessments for Certification and Licensure Decisions. *Simul Healthc* 2009; 4(1):35-42.
36. Integrated Comprehensive Clinical and Communication Skills Approach For Patient Interviews & OSCE Exams: What is the OSCE Examiners' Checklist? Available from URL: http://www.oscehome.com/OSCEs_Examiner_Checklist.html
37. Gorter S, Rethans JJ, Scherpbier A, van der Heijde D, Houben H, van der Vleuten C, et al. Developing case-specific checklists for standardized-patient-based assessments in internal medicine: a review of the literature. *Acad Med* 2000; 75(11): 1130-7.
38. Chumley HS. What Does an OSCE Checklist Measure? *Fam Med* 2008; 40(8): 589-91.
39. Cusimano MD, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R. A comparative analysis of the costs of administration of an OSCE. *Acad Med* 1994; 69(7): 571-6.
40. Association of Standardized Patient Educators. Available from URL: <http://www.aspeducators.org/>
41. Society for Simulation in Healthcare. Available from: URL: <http://www.ssih.org>
42. Hardee JT, Kasper IK. From Standardized Patient to Care Actor: Evolution of a Teaching Methodology. *Perm J* 2005; 9 (3):79-82.
43. Howley LD, Simons DF, Murray JA. Focusing Feedback on Interpersonal Skills: A workshop for Standardized Patients. 3rd edition. Unpublished training manual; 2005.
44. Sinclair N. Giving verbal feedback training for standardized patients. Presented at Annual Meeting ASPE; 2003.
45. Health Sciences Academic Services and Facilities. Clinical Skills and Laboratory Services. Training Your Standardized Patient; Training Your Standardized Patient to Give Feedback. Standardized Patient Trainer Information and Workshop. Available from URL: <http://deptswashington.edu>
46. Sinclair N. Reflective verbal feedback: A substrate for Professionalism (Workshop). Third annual meeting of the association of standardized patient educators. September 18-22, 2004, New Orleans, LA USA, Hosted by Tulane University School of Medicine; 2004.
47. Nelles LJ, Knickle K, McNaughton N, Tabak D. Beyond Sandwich-Advanced Feedback Skills (Workshop), 3rd Annual meeting of the association of standardized patient educators. September 18-22, 2004, New Orleans, LA USA, Hosted by Tulane University School of Medicine; 2004.

48. Adamo G, Brownfield E, Durning S. Objective Structured Clinical Examinations and Standardized Patients in Medical Education: Getting Started and Expanding roles. Presented at 2003 CDIM national meeting, Savannah GA; 2003.
49. Improving verbal skills. [Serial online] 1997 Aug; Available from: URL: <http://www.itstime.com/aug97/htm>
50. Hatchett P, Haun C, Goldenhar L. Training Standardized Patients to Give Feedback to Medical Trainees: The State of Art. University of Cincinnati College of Medicine, Department of Education, Center for Clinical Development and Assessment. Available from: URL: http://www.aspeducators.org/Hatchett_Feedback.pdf.
51. Cox K. No Oscar for OSCE. *Med Educ* 1990; 24(6): 540-5.
52. Leichner P, Sisler GC, Harper D. A study of the reliability of the clinical oral examination in psychiatry. *Can J psychiatry* 1984; 29(5): 394-397.
53. Norman GR, Barrows HS, Gliva G, Woodward. C. Simulated patients. In: Neufield VR, Norman GR, editors. *Assessing Clinical Competence*. New York: Springer; 1985. P. 219-29.
54. Sanson-Fisher RW, Poole AD. Simulated patients and the assessment of medical students' interpersonal skills. *Med Educ* 1980; 14(4): 249-53.
55. Thistlethwaite J, Silverman J, Ridgway G. A practical guide to working with simulated patients and as a simulated patient. Making it real. Radcliffe Publishing; 2006.
56. Barrows HS. *Simulated Patients: The Development and Use of a New Technique in Medical Education*. Springfield, IL: Charles C. Thomas; 1971.
57. Collins JP, Harden RM. The Use of Real Patients, Simulated Patients and Simulators in Clinical Examinations 2004. Association for Medical Education in Europe (AMEE) Guide No 13 2004; Available from URL: <http://www.medev.ac.uk/resources/features/>
58. 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(9):993-7.
59. Norman GR, Muzzin LJ, Williams RG, Swanson DB. Simulation in health sciences education. *J Instructional Dev* 1985; 8(1): 11-17.
60. Keiko A, Kei M, Tomio S, Nobutaro B. Shame is barrier to standardized patient's participation in the physical examination, but it can be mitigated: findings from a national-wide survey in Japan. *Nagoya J Med Sci* 2007; 69(1-2):86-7.

Mehran Zarghami, MD

Professor of Psychiatry, "Department of Psychiatry" and "Psychiatry and Behavioral Sciences Research Center", Mazandaran University of Medical Sciences and Health Services

• **Corresponding author:** Mehran Zarghami MD, Zare Hospital, Sari, Mazandaran, I.R.Iran
Tel: +98 151 32 85659
Fax: +98 151 3285109
Email: mzarghami@mazums.ac.ir