EMPIRICAL REPORT

Assessing Medical Students' Communication Skills by the Use of Standardized Patients: Emphasizing Standardized Patients' Quality Assurance

**M. Shirazi · A. Labaf · F. Monjazebi · M. Jalili ·
M. Mirzazadeh · S. Ponzer · I. Masiello**

Abstract

Objective The objective structured examination is one of the most valid, reliable, and effective tools for assessing clinical and communication skills, often by use of standard patients (SPs). SPs can also be assessors of those skills. One of the crucial areas when utilizing SP-based assessment is the quality and consistency assurance of their portrayal of the case and their ability to fill in checklists in an adequate way. The aim of this study was to assess the validity and reliability of SPs' ability to assess students' communication skill via a Calgary-Cambridge checklist.

Method This cross-sectional and correlational study was conducted at the Tehran University of Medical Science. We first analyzed *validity*; the criterion validity of the SPs' filling in the checklists was assessed through determining the correlation between the SPs' completed checklists and the checklists filled in by three physician raters individually and then *reproducibility:* it was assessed by a test-retest approach *inter-rater reliability.*

Result The mean correlation for assessing the validity of SPs' completed checklists by individual SPs was 0.81. The inter-rater reliability was calculated by kappa coefficient, and the total correlation among the three raters was 0.85. The reliability of the test-retest approach showed no significant differences between the test and re-test results.

Conclusion The increased number of medical students and different faculties' responsibilities such as doing educational, research, and health services duties assessing medical student communication skills is a complex issue. The results of our study showed that trained SPs can be used as a valid tool to assess medical students' communication skills, which is also more cost effective and reduces work load of medical faculties.

Physicians with good communication skills can easily extract a proper history from a patient, formulate an appropriate diagnosis, develop a doctor-patient relationship, and discuss strategies regarding patient management [1, 2]. Communication skills are the essential and also the main competency not only for general practitioners but also for psychiatry residents and psychiatrists. It is the primary ability of a psychiatrist in order to elicit the patient's main problem during history taking [3].

Evidence shows that communication skills can be learned [4], but such skills are not easily assessed by traditional methods, such as written exams [5]. Today, the objective structured clinical examination (OSCE) is one of the most valid, reliable, and effective tools for assessing clinical and communication skills [6]. One of the important components of an OSCE is the use of standardized patients (SPs), especially for the assessment of health professionals' communication and clinical skills performance [6, 7]. SPs are individuals trained to act as real patients by simulating a set of symptoms. Ever since SPs were introduced in the 1960s by Howard Barrows, their use has increased in the field of medical education, both for training and for assessment of students' competences [8].

An advantage in utilizing SPs is that it can provide adequate feedback on the students' performance after each

M. Shirazi · A. Labaf · F. Monjazebi (✉) · M. Jalili ·
M. Mirzazadeh
Tehran University of Medical Science, Tehran, Iran
e-mail: fmonjazebi@gmail.com

M. Shirazi · S. Ponzer · I. Masiello
Karolinska Institutet, Stockholm, Sweden

encounter, which has been consistently shown to be effective for improving performance [9]. Research on SPs has demonstrated that well-trained SPs not only can effectively and convincingly imitate medical conditions but they can also perform in a remarkably consistent way with high inter-rater agreement [10, 11]. According to Norman [12], by using well-trained SPs under standardized conditions, it is possible to assess several students and reduce the variability of the assessments, thereby providing an equivalent and fair examination for all students. In addition, it has been shown that when SPs are used under controlled and standardized conditions, they will perform consistently [13].

The use of SPs in training situations is also preferred in comparison with real patients for ethical and patient safety issues [13–16]. The individual SPs can be used as assessors of medical students' performance. However, a crucial area is the quality and assurance of consistency in their portrayal of the case and their ability to SPs fill out checklists adequately. Shirazi et al. [17] have previously examined a series of SP qualities in the field of assessment of depression disorder, i.e., the reliability of SPs' portrayal via a test-retest approach, the reliability of the use of a checklist by determining inter-rater reliability between groups of SPs and finally, examination of validity by filling out an observational rating scale. This study was done by three independent psychiatrist faculties, who watched and scored videotapes of SPs' performance. Based on the evidence, the validity of an SP-based assessment can be demonstrated by the variation in the test scores, which is indicative of actual variation in the examinees' clinical competence. Thus, a generally accepted indicator of clinical competence is needed as a gold standard criterion to validate SP-based assessments and to guide scoring in a standard setting. One recurring suggestion for the gold standard criterion is global ratings by faculty-physician observers [18].

In one published study, a panel of five faculty physicians observed and rated videotaped performances of 44 medical students on the seven-case New York City Consortium SP-based assessment [19]. Correlations between the scores on the actual examination and the faculty ratings ranged from 0.60 to 0.70, which were high enough to suggest that they could be improved by increasing training sessions for SPs [19].

Use of SPs for assessing clinical performance is widespread. Most studies focus on accurate portrayals of case specifics, usually a set of facts concerning symptoms and medical history. However, only a limited number of studies, especially in Eastern countries, evaluated SPs' reliability and validity as assessors when using checklists. There might be cultural differences in comparison to the Western countries that might be of interest. Most published studies focus on monitoring of SP portrayal accuracy, and some of them focus on the SPs filling out checklists in real workplaces [17]. In addition, various validated tools (checklists) have been used to assess medical students' communication skills, e.g., SEGUE,

Kalamazo, Common Ground, and Calgary-Cambridge (CC). Even though CC is well known, it has not been used to validate how SPs fill out checklists for assessing medical student communication skills in Eastern countries.

The aim of this study was to assess the inter- and intra-rater reliability, as well as the validity, of SPs' assessments of medical students' performance by means of the CC checklist. CC was chosen for the current study because it had been used for assessing medical students' communication skills in OSCE examinations at the Tehran University of Medical Science (TUMS) (unpublished data). The SPs were trained specifically for this study to ensure the standardization of the cases presented.

Methods

This cross-sectional and correlational study was conducted between November and December 2010 at (TUMS).

Participants

Announcements inviting people of various ages (with an emphasis on older age groups) were distributed near public places. Fifteen respondents to the invitation were interviewed and assessed using certain criteria related to well-being, availability, age, educational level, gender, and importance of payment. Twelve individuals were assessed as eligible, although only ten decided to participate in the training. SPs and medical students participated in this study. Ten SPs were recruited, between 21 and 63 years of age, and four were men. They were retired teachers and students.

In total, communication skills of 30 medical students in the fourth year of their study were assessed by SPs. The goal of the assessment was to determine the accuracy of SPs' competency in filling out CC checklists. The participating students were given a medical dictionary as an incentive for cooperation in the project.

On the basis of previous research [17], the following components of an SP program should be considered when testing its validity and reliability:

- Content (scenario and measures)
- Process (SPs' portrayal and SPs' ability to fill out checklists)

Content

Compiling Scenarios In this study, a scenario was based on a patient with stomach pain, developed by a group of experts consisting of a medical educationalist, internist, and two physicians who specialized in emergency medicine. The emphasis

was put on communication between the physician and the patient. The scenario they compiled was based on the main objectives of the CC guidelines [19, 20]. To facilitate the case-writing process, a case template was provided in which the experts could fill out information regarding symptoms, past medical history, family history, findings on physical examination, and so on[20]. The content validity of the scenarios was determined by consensus of an expert panel of ten faculty members from the Medical Education and Internal Medicine departments at TUMS. Optimally, written SP scenarios should be highly detailed, envisaging more information than any GP may elicit. To be convincing, the SP should respond with the same certainty as a true patient might show when answering any of the questions.

Measures—Observational Rating Scale The performance of the SPs was assessed by using the previously validated observational rating scale [16, 17] comprising five items for verbal and four for nonverbal communication. The scale was developed for the purpose of this study, and the content validity of the scale was determined by consensus of the expert panel.

Calgary-Cambridge (CC) Checklist The CC checklist was chosen as the tool for assessing communication skills in the study because it was already in use at TUMS Medical School. The validity and reliability of the CC checklist have been demonstrated in a previous study in Iran (unpublished data), as well as in other parts of the world [9, 19–22]. Four domains of the CC checklist are related to communication skills: interviewing and collecting information, counseling and delivering information, personal manners, and reporting. The questionnaire consists of seven parts: introduction, information gathering, assessing the patients' perception, structured interview, building a relationship, explanation, and management planning. There were in total 27 questions, and each question was to be answered on a Likert scale (0–2), where a score of 0 indicated an overall poor performance and a score of 2 an excellent performance of a student.

Case-Specific Assessment The case-specific assessment (CSA) was an SP-based performance examination that required medical students to demonstrate their clinical skills in a simulated medical environment. They had 15 min to interact with each of the SPs and 10 min to document and interpret their findings after the encounters. History taking (Hx) and communication skills were assessed via the CSA scored by the SPs following the encounters.

Process

The training process was based on the Peggy Wallas book "Coaching Standardized Patients: for Use in the Assessment of Clinical Competence" [8]. The author

emphasized the importance of six separate items for both the SPs and the coaches in order to optimize an SP performance:

1. Realistic portrayal of the patient
2. Appropriate and correct responses to what students ask or do
3. Precise observation of the medical student's behavior
4. Rigorous recall of the student's behavior
5. Accurate completion of the checklist
6. Effective feedback to the student (written or verbal) on how the patient experienced the interaction with the student

The recruited SPs were trained by three coaches (i.e., one emergency medicine physician, one medical educationalist, and one psychiatrist) in a small group setting with five educational sessions of 2 h each. The training focused on the SP portrait. An additional 5-h training session was provided on how to fill out the checklists concerning the medical students' performance during the OSCE. The educational session had two phases. During the first phase, including three educational meetings, the SPs played their role with each other under the supervision of their coaches and received feedback on their role-playing. In the second phase of training (five sessions), they learned how to fill out the CC checklists. The coaches played the roles of medical students and asked the SPs to rate their performance using the CC checklists. Then, the SPs discussed the results of their completed checklists and were given appropriate feedback by the coaches. All the sessions were video-recorded, and the videos were given to the SPs for further training in their roles.

Educational Material The printed material consisted of scenarios and detailed information on the communication skills according to the CC checklists and guidelines. The SPs read the handouts and watched videos regarding communication between a patient and a physician in order to better understand doctor-patient relationship and thereby get a clear idea about how he or she should portray the patient role and how to answer the questions on the checklists.

Validation of SP Process Although the main aim of this study was to utilize SPs as a valid instrument for assessing medical students' performance in communication skills, we encouraged the SPs to accurately portray the role and fill out the checklist.

Validation of SP Portrayal For the performance in their portrayals to be rated as indistinguishable from that of real patients, the SPs had to have better than 90% overall accuracy rate in the content of their presentations. The SPs' performance was assessed by three experts using an observational rating scale.

Validation of SPs' Ability to Fill Out the Checklists One week after the end of training courses, the SPs role played with three medical students for the first time. Then, the correlation between the SP completed checklists and three experts' judgments (as a gold standard) were investigated for each SP separately. Each of the ten SPs' encounters with the three medical students was video-recorded. One week following the first encounter, each SP met one of the same medical students for the second time.

Validation of SPs' Filled Out Checklists

Validity The criterion validity of the SPs' completed checklists was assessed by determining the correlation between the SPs' completed checklists and the checklists filled out by the three raters individually.

Reproducibility The reproducibility was assessed using a test-retest approach, where the SPs' initial checklists were compared with the checklist completed one week following their first encounter with the medical students.

Inter-rater Reliability The inter-rater reliability was tested by assessing the correlation between raters and applying non-parametric analysis tests [22, 23] .

Ethical Considerations

The Ethics Committee at TUMS approved this study. The SPs were informed that they would remain anonymous in all publications of the results and that they could receive direct support from the main investigator (MSh) if they encountered any problems. The students who performed with the SPs were required to sign an informed consent form.

Data Analysis

A data analysis was performed using SPSS software version 16. The validity was assessed by Spearman's rho test for correlation. The reliability of the test-retest approach in the SPs' ability to fill out the CC checklist was assessed by means of a Student's *t* test analysis. The inter-rater reliability was assessed by use of the kappa coefficient.

Results

The mean age of the SPs was 41 years (SD±16.8); 73% of them were women and 50% were married. They were retired teachers, students, and housewives. The medical students' age range was 22–26 years. They were enrolled in their fourth year and 70% of them were women. The SPs' performance is provided in Table 1 and 2. The mean correlation on assessing the validity of the SPs' completed individual checklists was 0.81 (range: 0.5 to 1) (Table 1). The checklists' reliability, Cronbach's alpha, was calculated to be 0.76. The inter-rater reliability kappa coefficient between rater 1 and rater 2 was 0.70 ($P=0.000$), between rater 1 and rater 3, 0.80 ($P=0.000$), and between rater 2 and rater 3, 0.60 ($P=0.001$). The total correlation between the three raters, the intraclass coefficient, was 0.85. The results showed no significant differences between the test and re-test results (Table 2).

Discussion

Several studies have shown that the interaction that occurs in medical encounters is remarkably influenced by the doctor's competence in communication skills [24]. Improving doctors' communication skills competency in different specialties is recommended by most of the acceptable health care societies such as Accreditation Council for Graduate Medical Education (ACGME)[25] .

Silverman, who developed the Calgary-Cambridge guidelines for medical interviewing at Cambridge University, claimed that communication skills are one of the core elements of the medical curriculum. Cambridge University has integrated communication skill training in all parts of the curriculum [26]. Generally, communication skill is provided as a separate course in most parts of the world, like in University Medical Center Hamburg-Eppendorf in Germany [27] and TUMS in Iran. Meanwhile, in the integrated communication skills curriculum at Cambridge University, high standard assessment of communication skills through the OSCE examination is a very significant issue; passing or failing in the communication skill stations is equal to a failure in the whole exam without considering other assessment components (e.g., multiple choice questions, short essay questions, and OSCE on physical exam). Hence, it is an extremely high-stakes examination [26]. Holding high-stakes exams will lead to patient safety, providing that quality assurance of the SP program and considering SP feedback (from a patient perspective) to students in OSCE communication skill stations are inevitable.

The certainty of the doctors' communication skills also underpins the goal of performance-based assessment programs. A high-stakes examination should be based on a valid interface between the SPs and the medical students to be assessed [28]. Therefore, the need for SPs to play their roles in a consistent fashion is of great importance. In this paper, we report the results of a cross-sectional study on the assurance of the SP-based assessment's quality in an OSCE setting. The results of this study showed that SPs are valid and reliable assessors of medical students' communication skill in an OSCE setting.

Table 1 The criterion validity of SP filled-out checklists (ten SPs) shown as the correlation (Spearman's rho) between the SPs and the three raters' scores

Rater SP	1		2		3		Mean of three raters	
SP1	r=0.82	P=0.41	r=1	P=0.01	r=1	P=0.01	r=0.94	P=0.01
SP2	r=0.51	P=0.66	r=1	P=0.01	r=0.50	P=0.67	r=0.70	P=0.66
SP3	r=0.87	P=0.33	r=1	P=0.01	r=0.86	P=0.30	r=0.91	P=0.33
SP4	r=1	P=0.01	r=1	P=0.01	r=1	P=0.01	r=1	P=0.01
SP5	r=0.65	P=0.54	r=0.50	P=0.66	r=0.67	P=0.51	r=0.60	P=0.67
SP6	r=1	P=0.01	r=1	P=0.01	r=1	P=0.01	r=1	P=0.01
SP7	r=0.51	P=0.66	r=0.50	P=0.66	r=0.50	P=0.66	r=0.55	P=0.67
SP8	r=1	P=0.01	r=1	P=0.01	r=1	P=0.34	r=1	P=0.33
SP9	r=0.86	P=0.33	r=0.86	P=0.33	r=0.50	P=0.66	r=0.74	P=0.33
SP10	r=1	P=0.01	r=1	P=0.01	r=1	P=0.01	r=1	P=0.01

Using SPs as raters for assessing students' performance in an OSCE is valuable because it avoids the probable bias of faculty who might be affected by their earlier perceptions and knowledge of the students [29]. Besides, clinical faculties are often too busy to spend a lot of time in evaluating these skills.

Our main finding is in line with previous studies in different fields of medicine in the Western countries and emphasizes the fact that a trained SP can perform as a reliable rater for evaluating doctors' behaviors, including communication skills.

Quality Assurance of SP Consistency

SP standardization is necessary if acceptable SP-based assessment conditions are met, but it is a challenging task for anyone employing SPs for assessments [30]. One of the main issues emphasized when using SPs in examinations is the consistency of the SPs, which is part of the standardization process. The

Table 2 Test and re-test results (mean scores) of the checklists, filled out by the SPs

SP number	Test Mean score	Re-test Mean score	Student's t test	P value
1	42	42	1.87	0.10
2	48	50	0.62	0.55
3	48	50	1.52	0.17
4	38	40	1.87	0.10
5	45	50	1.93	0.09
6	46	49	2.04	0.08
7	36	42	2.12	0.07
8	49	50	1.52	0.17
9	43	47	0.10	0.92
10	49	48	1.52	0.17
Total	44.40	46.80		
Standard deviation	4.600	3.938		

assessment of SPs' performance is based on the authenticity and accuracy (consistency) of their presentation. Both of these issues were examined in this study using different validity and reliability measures, i.e., alpha, kappa, and inter-class values, as follows.

Validity of the SP Assessment

Validity refers to how well an SP is trained and plays his/her role correctly, and whether a checklist is used to ensure that the training is standardized. It is important for the simulation to be recorded [31]. In this study, the criterion validity of the SPs' checklists completion was examined by correlating between the SPs' and the raters' scores. SP ratings predicted the students' competence in communication skills, and this was confirmed by a high correlation between the SPs' and the examiners'/raters' scores in each case. These results conform with an earlier study by Zanten et al. [24], who found a high correlation between examiners' and SPs' scores in an international physician clinical skills assessment in an American medical setting . Rothman and colleagues also found similar results, although they noted less agreement [32].

Whelan et al. [33] recommend an integrated form of assessment. They suggest that communication skills could be assessed by SPs and problem solving skills by physicians. In contrast, other researchers have found weak correlation between SP-based assessment scores and physicians' scores and argue that physicians, not SPs, are the only qualified experts who should judge performance [32]. Nonetheless, a small sample size or issues with SPs' training and standardization can affect the results [7, 8]. It is important for the SP to complete, within the encounter time frame, all the checklists with better than 85% overall accuracy rate and give effective written feedback. Validity requires a degree of consensus among experts regarding the key features that should be included in a given case. Hence, assessment tools and other training protocols should be chosen and validated by more

than one expert to ensure that all key features are included [34]. These issues were considered in our present study, and the educational protocol and content (case and tools) were developed to increase the criterion validity.

Inter-rater Reliability

Our results demonstrated a strong correlation between the individual raters (kappa coefficient) and a good correlation between all three raters (intra-class correlation coefficient). In line with our findings, another study has also found acceptable or highly acceptable coefficients [18, 19]. In contrast, yet another study has found weak correlation between its raters. The latter finding may be due to the fact that the authors did not follow the guidelines and did not organize an expert panel to develop checklists and cases [20, 30]. To increase inter-rater reliability between SP assessors in this study, we compiled a guideline for the assessment and based the filling out of the checklists on the scenario. Moreover, the SPs and the medical students were trained before they started to rate the videos individually. As stated above, the reproducibility of the SPs' completed checklists was tested by a paired t test analysis. The data showed that there were no significant differences between the test and the re-test scores. The SPs in our study had been trained to fill out the checklists in a consistent and accurate manner on several occasions and had also received feedback from their coaches, which may have helped the positive results.

Hence, utilizing an SP as an assessor in an OSCE is an equitable solution and may even lower the cost of the OSCE. A clinical faculty is often overloaded with numerous other duties and may be biased because of their previous familiarity with the students. However, it is not always possible to replace physicians with SPs in all parts of a clinical skills assessment. The validation of scores relies on the raters' accuracy; therefore, the process and content of SPs' training must be appropriate to ensure the quality of their rating performance. Systematic training techniques are necessary; for example, training on how to fill out valid and reliable checklists, clear scenarios, coaching for correct and standard role playing, and constructive feedback [32].

Quality Assurance of SP Content

One of the important components of this study was compiling standardized case scripts. Therefore, much time and effort was devoted to this task by the expert panelists. This procedure was also emphasized by Boulet et al. in 2009[30]. Our results showed that the Calgary-Cambridge checklist had high internal consistency and reliability, which were in accord with previous studies [22] .

Methodological Consideration

The novelty of the current study is not only for being the first conducted in the Middle East region but also by focusing on different cultural perspectives between the doctor and the patient in Eastern countries in comparison with Western countries. There is an expectation regarding the increased risk of the potential biases through SPs' scoring of medical students (SPs' higher scoring in comparison with faculties' scoring), which could be seen as a limitation of the study, due to the higher sociocultural situation of doctors in the eastern area. However, because of the remarkable SP trainings, our results demonstrated a strong correlation between the scoring of SPs and faculty members, which rule out the assumption regarding the effect of cultural diversity on SP scoring.

Our study had a small sample size, which might have affected the results. Moreover, we trained the SPs on the basis of only one standardized case, which may lower the generalizability of the results.

Additional studies will be needed to identify the sources of case variability. Of particular importance is the investigation of the relationship between the domain of knowledge and the quality of communication [30, 34] .

The increasing number of medical students combined with faculty members' responsibilities for education, research, and providing health services, as well as the assessment of medical students' communication skills, combine to make assessments more difficult. The results of our study showed that trained SPs can be used as a valid tool to assess medical students' communication skills. This will build the case for employing SPs in more ways, not only can they help in the evaluation of topic-specific medical competence of students but can also be useful in the separate domains of communication skills, which are also more cost effective and might reduce the work load of medical faculties.

References

1. Baig LA, Violato C, Crutcher RA. Assessing clinical communication skills in physicians: are the skills context specific or generalizable. BMC Med Educ. 2009;9:22.
2. Maguire P, Pitceathly C. Key communication skills and how to acquire them. Brit Med J. 2002;325:697–700.
3. Stubbe D. Effective patient centered—communication FOCUS. Winter. 2012;x(1):42–4.
4. World Health Organization. Doctor-patient interaction and communication. Geneva: WHO; 1993.
5. Vu NV, Barrows HS. Use of standardized patients in clinical assessments: recent developments and measurement findings. Educational Researcher 1994;23(3):23–30.

6. Adamo G. Simulated and standardized patients in OSCE achievement and challenges. Med Teach. 2003;25:262–70.

7. Hodges B et al. Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. Med Educ. 1996;30:38–43.

8. Wallace P. Coaching standardized patients: for use in the assessment of clinical competence. In: Wallace P, editor. New York: Springer Publishing Company; 2006.

9. Schlegel C et al. Validity evidence and reliability of a simulated patient feedback instrument. BMC Med Educ. 2012;12:1–6.

10. Thompson JC, Kosmorsky GS, Ellis BD. Field of dreamers and dreamed-up fields: functional and fake perimetry. Ophthalmology. 1996;103:11725.

11. Colliver JA, Swartz MH. Assessing clinical performance with standardized patients. JAMA. 1997;278:7901.

12. Norman GR et al. Simulated patients. In: Neufeld VR, Norman GR, editors. Assessing clinical competence. New York: Springer Publishing Company; 1985.

13. Gomez JM, et al. Clinical skills assessment with standardized patients. Med Educ. 1997;31(2):94–8.

14. Ponzer S, Hylin U. Standardized patients—good help in teaching. They function as stand-ins in training of consultation and clinical skills. Lakartidningen. 2004;101(42):3240–4.

15. Vu NV. Steward DE, and Marcy M, An assessment of the consistency and accuracy of standardized patients' simulations. J Med Educ. 1987;62:1002.

16. Colliver JA SM, Robbs RS, Lofquist M, Cohen D, Verhulst SJ. The effect of using multiple standardized patients on the intercase reliability of a large scale standardized patient examination administered over an extended testing period. Acad Med. 1998;73(10 suppl): S813.

17. Shirazi M et al. Training and validation of standardized patients for unannounced assessment of physicians' management of depression. Acad Psychiatr. 2011;35(6):382–7.

18. Colliver JA. Validation of standardized-patient assessment: a meaning for clinical competence. Acad Med. 1995;70:1062–4.

19. Swartz MH, Colliver JA. Using standardized patients for assessing clinical performance: an overview. Mt Sinai J Med. 1996;63:241–9.

20. Chan CSY et al. Communication skill of general practitioners: any room for improvement? How much can it be improved? Med Educ. 2003;37:514–26.

21. Whelan R et al. Scoring standardized patient examinations: lessons learned from the development and administration of the ECFMG Clinical Skills Assessment (CSA®). Med Teach. 2005;27:200–6.

22. Donnon T, Lee M, Cairncross S. Using item analysis to assess objectively the quality of the Calgary-Cambridge OSCE Checklist. CMEJ, 2011;2(1):e16–e22.

23. Keen AJA, Klein S, Alexander DA. Assessing the communication skills of doctors in training: reliability and sources of error. Adv Health Sci Educ. 2003;8:5–16.

24. Zanten VM, Boulet JR, McKinley D. Using standardized patients to assess the interpersonal skills of physicians: six years' experience with a high-stakes certification examination. Health Commun. 2007;22(3):195–205.

25. Wasnick JD et al. Do residency applicants know what the ACGME core competencies are? One program's experience. Acad Med. 2010;85(5):791–3.

26. Silverman J. Teaching clinical communication: a mainstream activity or just a minority sport? Patient Educ Couns. 2009;76(3):361–7.

27. Hausberg MC et al. Enhancing medical students' communication skills: development and evaluation of an undergraduate training program. BMC Med Educ. 2012;12(1):16.

28. Boulet JR. Using SP-based assessments to make valid inferences concerning student abilities: pitfalls and challenges. Med Teach. 2012;34(9):681–2.

29. Brinkman WB et al. Evaluation of resident communication skills and professionalism: a matter of perspective? Pediatrics. 2006;118(4):1371–9.

30. Boulet JR et al. Checklist content on a standardized patient assessment: an ex post facto review. Adv Health Sci Educ. 2008;13:59–69.

31. Fincher R-ME, Lewis LA. Simulations used to teach clinical skills. In: Norman GR, van der Vleuten CPM, Newble DI, editors. Int Handb Res Med Educ. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2002. p. 499–535.

32. Parkes J, Sinclair N, McCarty T. Appropriate expertise and training for standardized patient assessment examiners. Acad Psychiatr. 2009;33(4):285–8.

33. Rothman AI, Cusimano M. A comparison of physician examiners', standardized patients', and communication experts' ratings of international medical graduates' English proficiency. Acad Med. 2000;75(12):1206–11.

34. Gutton G et al. Communication skills in standardized-patient assessment of final-year medical students: a psychometric study. Adv Health Sci Educ. 2004;9:179–87.