

Comparison of Assessment Scores of Candidates for Communication Skills in an OSCE, by Examiners, Candidates and Simulated Patients

Abdul Sattar Khan¹, Riaz Qureshi², Hamit Acemoğlu³, Syed Shabi-ul-Hassan⁴

¹Department of Family Medicine, Ataturk University, Erzurum, Turkey

²Department of Family Medicine, King Saud University, Riyadh, Saudi Arabia

³Department of Medical Education, Ataturk University, Erzurum, Turkey

⁴Department of Family Medicine, Riyadh Military Hospital, Riyadh, Saudi Arabia

Email: yardockhan.ask@gmail.com, abdulstatar@atauni.edu.tr

Received July 12th, 2012; revised August 10th, 2012; accepted August 24th, 2012

Though OSCE method has been verified by several researchers for the appropriate assessment of competence in clinical skills, yet medical educationists have some concerns regarding the value of assessment of communication skills and empathy by this method. Hence, we sought to assess the extent of differences, if any, among the examiners, the candidates and the simulated patients (SPs) for communication skills. A total of 23 general practitioners, who were preparing for their postgraduate clinical examination, participated in a practice OSCE on seven stations in this study. The examiners observed and evaluated the candidates during the whole consultation, using the pre-tested checklist including 15 items with a global rating scale. The simulated patients also evaluated the candidates at the end of consultation, using the same checklist. There were significant differences in the assessment scores of candidates by the examiners, the candidates themselves and the simulated patients regarding all aspects of communication skills. However, introduction to the patient's scenario of some non-verbal communication did not show any significant difference (p -value = 0.05). The correlation between examiners and SPs ($r = 0.07$, $p = 0.7$) and SPs and candidates ($r = 0.01$, $p = 0.95$) was very low and not significant. Cronbach's alpha was 0.968 across items, whereas among seven stations it was 0.931. This study has shown a significance difference in assessment scores of candidates by examiners, SPs and candidates themselves. In conclusion, there is a need for further research regarding the active role of SPs in summative assessments.

Keywords: Assessment by Simulated Patients; Empathy; Communication Skills; OSCE

Introduction

The objective structured clinical examination (OSCE) was pioneered in medicine in the late 70s as a tool for ensuring standardization and psychometric stability in high-stakes assessments of clinical skills (Harden & Gleeson, 1979). This method has been discovered to add to the ward-based teaching and the recognition that students require more opportunities to practice in a controlled environment, prior to actually being released in a clinical setting (Harden & Gleeson, 1979; Robb, 1985). Professional actors have been trained to portray as patients and this practice has become a commonplace in many health professions assessments (Bokken, van Dalen, & Rethans, 2010; Watson, 2004). Self-assessment of knowledge and accuracy of performance of clinical skills is essential to the practice of medicine and self-directed life-long learning (Pierre RB). Recently, senior students are also being used as an examiner to support faculty members, especially for formative assessments (Moineau, Power, Pion, Wood, & Humphrey-Murto, 2011).

The third main component of the whole process is the patient; although the importance of feedback by simulated patients (SPs) is generally recognized to be useful, knowledge is scarce about the most effective way in which SPs can provide feedback. In addition, little is known about how SPs are trained to provide feedback (Bokken, Linssen, Scherpier, van der Vleuten, &

Rethans, 2009) and further, whether there is any role of their input in the assessment during an OSCE (Thistlethwaite, 2002). Physician-patient communication including empathy, which is a highly complex process, can be tested by OSCE according to some studies (Fischbeck, Mauch, Leschnik, Beutel, & Laubach, 2011). However, reliability of the global scoring by examiners as observers is still debatable (Schwartzman, Hsu, Law, & Chung, 2011b). Additionally, this part of consultation is purely related to understanding of patients or simulated patients (SPs), that might be difficult to understand only by observation, without taking opinion of patients. A recent systemic review also emphasized this point (Brannick, Erol-Korkmaz, & Prewett, 2011). It raises several questions, for example how, when and where to get SPs opinion and whether it adds any valuable results to assessment of communication skills (Rosen, 2008).

Perhaps, addressing to these issues requires an objective evaluation to understand the role of simulated patients (SPs) in assessment of communication skills and to look at any differences among examiners (as observers) assessment, candidates' self-assessment and SPs assessment of the same station. Within this context, we attempted to find out, whether there is any significant difference in the assessment of performance of the candidates, among examiners, candidates themselves and the simulated patients and, does it have an effect on the overall

results of OSCE as regards to the evaluation of communication skills, by using a global rating scale.

Method

Study Design

This was a descriptive exploratory study, mainly focused on the general practitioners (GPs), who participated in a training MOCK examination for the preparation of an international postgraduate examination. The candidates had completed the scheduled and mandatory clinical skills training with clinical faculty during their preparatory course. The SPs, who participated in this study, were a mix of junior doctors and nurses, who were trained to play the role of SPs in several, previous mock OSCEs. There were 07 stations in the OSCE, which comprised of stations with a focus on history taking and communication/counseling skills and excluded physical examination. The topics of these stations consisted of: history of flank pain; counseling for oral contraceptives; post MI counseling; mild depression; counseling of mother of an obese child; explanation and discussion on PSA results and a case of menopause.

Study Setting

The study took place at a postgraduate training center, under Ministry of Health, Saudi Arabia during 2010. Examiners, SPs and Candidates were given a briefing session before the OSCE, where the goals and objectives of the study were also explained; queries and concerns were addressed and consent for participation was taken.

Instrument and Data Collection

A rating scale, consisting of 15 items relevant to specific history-taking and communication skills including some components of empathy was developed, keeping in view the objectives of previous communication skills training and literature (Allen, Heard, & Savidge, 1998; Chumley, 2008; Mazor, Ockene, Rogers, Carlin, & Quirk, 2005; Regehr, Freeman, Robb, Missiha, & Heisey, 1999). It was discussed with other senior faculty in order to check its face and content validity and was then applied to observe in real situation at a family medicine unit for pre-testing. An input was also taken from colleagues, whether they agreed with the items and rating scales or not.

The rating scale was developed based on literature (Hatala, Marr, Cuncic, & Bacchus, 2011; Schwartzman, Hsu, Law, & Chung, 2011a; Townsend, McIlvenny, Miller, & Dunn, 2001) and discussions with consultants of family medicine, psychiatry and medical education departments. The performance has assessed by using a ten-point response range. It consisted of not done, very poor, meager, marginal, satisfactory, good, very good, excellent, outstanding and exceptional. Satisfactory evaluation was the minimum passing criteria. The items included were: generic aspects of history taking, like questioning skills, professional manner and organization of interviews, with time management and closing of interviews, understanding of patients, discussion about patient's ideas, concerns and expectations, shared decision making with some non-verbal skills, like nodding head, good listener, eye to eye contact, leaning forward etc.

Data Analysis

Results were analyzed using SPSS 18.0 for Windows. For each attribute, mean and standard deviation of assessment scores for three groups: examiners, candidates and SPs were calculated. These were tested using ANOVA for significant differences with test of homogeneity. PostHoc test was performed later, to compare means within the groups and with the groups. Level of significance was set at $p < 0.05$. A Pearson chi-square test was applied for categorical data and correlations were also assessed among examiners, candidates and the SPs by calculations, using a Spearman rank order correlation. Test of reliability was applied to check Cronbach's alpha.

Results

There were 23 participating candidates, 12 females and 11 males. The seven examiners were well trained Family Physicians with postgraduate qualifications in Family Medicine. The reliability coefficient (Cronbach's alpha) showed 0.968 across items, whereas among seven stations it was 0.931. **Table 1** presents results of three assessor groups of competency in communication skills performance. The results depicted that among all three groups there is a significant difference ($p \leq 0.05$) in the assessment of performance of the candidates. The students rated themselves in almost all aspects of communication skills, above average level (mean score range from 5 to 9), while the assessment range by examiners (mean score range from 2 to 8) and by SPs (mean score range from 2 to 7) were somewhat similar.

Simulated patients assessment scores were below satisfactory level (< 5 mean score) in majority of the items in contrast to the examiners, who rated candidates above satisfactory (> 5 mean score) in majority of items in almost all stations (**Table 2**). In terms of overall performance, 56.5% of the candidates were declared to have achieved satisfactory or higher level scores (equal to 5 or more), by the examiners, whereas all the candidates rated themselves at satisfactory or higher level regarding their overall level of performance. The SPs on the other hand were less generous and rated that only 26% of the candidates performance was at satisfactory or above level.

Further additional exploration of the differences among means was needed to provide specific information on which means are significantly different from each other among examiners, candidates and SPs. Therefore post hoc ANOVA analysis was performed and the results are shown in **Table 3**. It has highlighted that out of 15 items evaluated by examiners and candidates, there is a significant difference in the performance of the candidates (< 0.05). On the contrary, examiners have shown no difference of opinion for five items, mainly related to non-verbal communication skills, as compared to the assessment by the SPs, like introduction to patients, patients understanding for explanation, good listener, eye to eye contact and leaning forward ($p = 0.05$).

The correlation between examiners and the candidates was moderate and significant ($r = 0.47, p = 0.023$), while between examiners and SPs ($r = 0.07, p = 0.7$) and SPs and candidates ($r = 0.01, p = 0.95$) it was very low and not significant (**Figure 1**).

Discussion

So far, much has been done to investigate the involvement of simulated patients (SPs) in medical training situations, with emphasis on clarifying the validity, standardization and

Table 1.
Comparison of mean scores given by examiners, candidates and SPs.

Items evaluated for communication skills	Observers Mean (SD)	Students Mean (SD)	Simulated Patients Mean (SD)	p-Value
Introduction to self to patient	5.26 (1.36)	7.26 (0.75)	4.78 (1.04)	0.001
Consent taken for history taking	5.22 (1.17)	6.87 (1.10)	4.39 (0.84)	0.002
Explore idea concern & expectations	5.17 (0.89)	7.00 (0.85)	4.22 (0.80)	0.001
Understand effect of problem on daily routine	5.22 (1.17)	6.87 (1.10)	4.22 (0.80)	0.004
Able to explain diagnosis	5.43 (1.12)	7.00 (0.85)	4.39 (0.84)	0.001
Communicate to patient about his/her concerns	5.22 (1.17)	7.26 (0.75)	4.22 (0.80)	0.001
Patients understand explanation (nodding head or responding verbally)	5.26 (1.36)	6.87 (1.10)	4.78 (1.04)	0.001
Offer help in a polite way	5.35 (0.98)	7.00 (0.85)	4.39 (0.84)	0.00
Share decision for management	5.23 (1.19)	6.87 (1.10)	4.22 (0.80)	0.003
Discuss safety netting	5.48 (1.08)	7.26 (0.75)	4.22 (0.80)	0.001
Nodding head	5.26 (1.38)	6.87 (1.10)	4.39 (0.84)	0.00
Good listener (Didn't interrupt patients)	4.78 (1.13)	7.00 (0.85)	4.22 (0.80)	0.005
Eye to eye contact	5.22 (1.17)	7.26 (0.75)	4.78 (1.04)	0.001
Lean forward	4.83 (1.59)	6.87 (1.10)	4.39 (0.84)	0.005
Shake hands	5.26 (1.36)	7.00 (0.85)	4.22 (0.80)	0.001

Table 2.
Overall results and mean score given by different assessors at OSCE stations.

Overall results by assessors	Pass N (%)	Fail N (%)
Observer	13 (56.5)	10 (43.5)
Examinee	23 (100)	00
Simulated Patients	06 (26.1)	17 (73.9)
Overall mean scores by assessors	Mean \pm SD (95% Confidence Interval)	
Observer	5.21 \pm 0.66 (4.92 - 5.50)	
Examinee	7.02 \pm 0.43 (6.83 - 7.20)	
Simulated Patients	4.39 \pm 0.58 (4.14 - 4.64)	

feasibility of the SP role as a teaching and assessment "tool" (Rosen, 2008); however, there has been less emphasis on evaluation of reliability of patient-centered approach, through assessment of communication skills. Furthermore, it is relatively difficult to reliably assess communication skills, as compared to clinical skills, when considering both as general traits, that should apply across multiple situations (Brannick et al., 2011).

Our study has demonstrated that, while assessing the role of SPs as assessors and comparing it with the assessment of examiners as observers, majority of items related to verbal communication have significant differences. However non-verbal domains did not show significant differences between the assessment by examiners and SPs. These findings second the

results published recently in a systemic review, which emphasized on the difficulty of measuring communication skills in a reliable manner.

Though, one can argue about the relatively small number of participants and stations in this study, but the findings of a wide range of difference in assessment of communication skills between the examiners and the SPs in the study cannot be totally ignored. Consequently it may alarm medical educationists, whether the candidates who pass their examination with dissatisfied simulated patients (SPs) would be able to practice as patient-centered physicians in real situation.

Several methods have been suggested to assess communication skills reliably and a recent systemic review (Brannick et al., 2011) also suggested using two examiners and large number of stations. As an argument, it is stated that better than average reliability is associated with a greater number of stations and a higher number of examiners per station. However, it sounds somewhat like a luxury and logistically difficult to implement in the light of the scarce resources in some developing countries.

In addition, when we talk about high-stakes examination, with large number of examinees, generally, stressful roles were indeed found to be stressful (McNaughton, Tiberius, & Hodges, 1999) and negative effects were said to be more evident when role players had complex situations to portray. McNaughton (McNaughton et al., 1999) suggested that in high-stakes psychiatric examinations, SPs had negative physical and emotional reactions that continued past the day of acting. As a result, observers or examiners in this situation, will not be able to appraise correctly and even SPs as assessors may give biased results, which will ultimately affect the candidates in terms of progress in career, learning process and moral or self-esteem

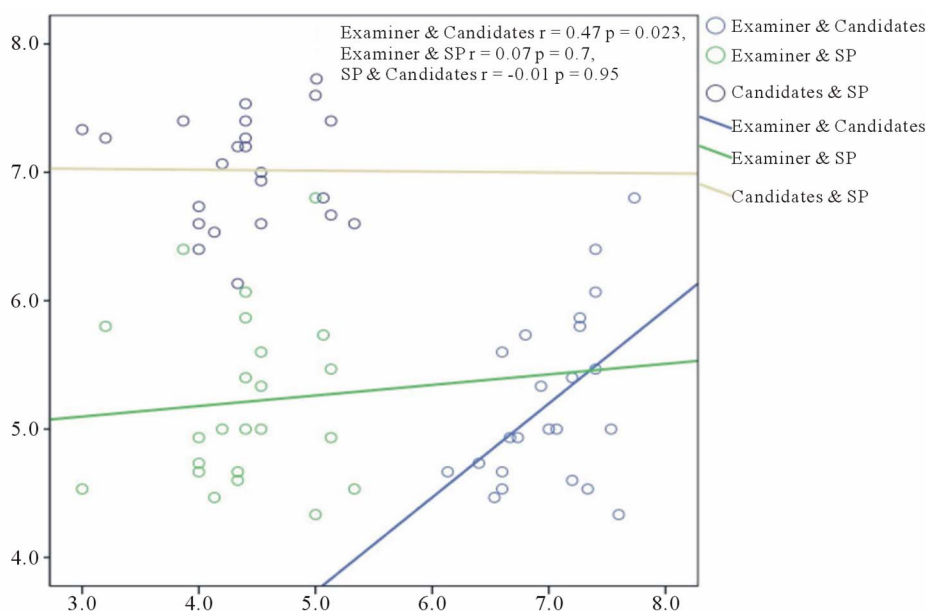


Figure 1.
Correlation among examiners, candidates and SP.

Table 3.
Comparison between groups of examiners and SP and candidates.

Items evaluated for communication skills	Group vs Group	Mean Difference	p-Value	95% CI	
				Lower	Upper
Introduction to patient	Obs	SP 0.47	0.329	-0.32	1.27
	Stud	-2.00	0.000	-2.80	-1.20
Consent taken for history taking	Obs	SP 0.82	0.033	0.05	1.60
	Stud	-1.65	0.000	-2.42	-0.88
Explore idea concern & Expectations	Obs	SP 0.951	0.001	0.35	1.57
	Stud	-1.82	0.000	-2.44	-1.21
Understand effect of problem on daily routine	Obs	SP 1.00	0.007	0.24	1.76
	Stud	-1.65	0.000	-2.42	-0.89
Able to explain diagnosis	Obs	SP 1.04	0.002	0.34	1.74
	Stud	-1.56	0.000	-2.26	-0.87
Communicate to patient about his/her concerns	Obs	SP 1.00	0.002	0.32	1.68
	Stud	-2.04	0.000	-2.73	-1.36
Patients understand explanation (nodding head or responding verbally)	Obs	SP 0.47	0.390	-0.39	1.35
	Stud	-1.60	0.000	-2.48	-0.74
Offer help in a polite way	Obs	SP 0.95	0.002	0.30	1.62
	Stud	-1.65	0.000	-2.31	-0.99
Share decision for management	Obs	SP 1.01	0.007	0.23	1.79
	Stud	-1.64	0.000	-2.42	-0.86
Discuss safety netting	Obs	SP 1.26	0.000	0.60	1.92
	Stud	-1.78	0.000	-2.44	-1.13
Nodding head	Obs	SP 0.87	0.037	0.04	1.70
	Stud	-1.60	0.000	-2.43	-0.78
Good listener (Didn't interrupt patients)	Obs	SP 0.56	0.131	-0.13	1.26
	Stud	-2.21	0.000	-2.91	-1.53
Eye to eye contact	Obs	SP 0.43	0.345	-0.31	1.17
	Stud	-2.04	0.000	-2.78	-1.30
Lean forward	Obs	SP 0.43	0.483	-0.46	1.33
	Stud	-2.47	0.000	-2.94	-1.15
Shake hand	Obs	SP 1.04	0.004	0.28	1.81
	Stud	-1.73	0.000	-2.50	-0.98

Note: Obs: Observer; SP: Simulated Patients; Stud: Students.

aspects. Of course, it is a difficult task to incorporate SPs as assessors in an OSCE and may not be feasible in terms of time management; however it is likely to be more reliable in assessing communication skills and could also be a cost saving exercise.

It has also been emphasized that candidates may be utilized in assessment process and self-assessment has already been established as a very effective learning tool, especially as regard to history taking, exploring presenting problems and taking drugs and family histories etc. (Regehr, G., 2006). Importantly, however there is always a problem of biased results. Yet interestingly, when we analyzed overall performance in our study, based on global scoring, the candidates rated themselves performance wise in 100% satisfactory or higher category, where as examiners assessed that a little higher than 50% candidates performed satisfactorily, and the SPs assessed that only one quarter of the candidates performed at or above satisfactory level.

The candidates on self-assessment rated their overall skills markedly higher than the assessment of their overall skills by the examiners and the SPs. This could be explained by the fact that physician-patient communication is a complex process and often has high subjectivity and may be influenced by task familiarity (Bianchi, Stobbe, & Eva, 2008; Taras, 2002). A few studies have shown that students tended to assess their skills much lower than expected by their teachers (Siaja, 2006); contrary to this, another study (Jahan, Sadaf, Bhanji, Naeem, & Qureshi, 2011) has shown comparable results as regard to communication skills. The results of our study do not match with these findings. One obvious explanation for these markedly different results could be due to the fact, that our small-scale study was conducted on experienced general practitioners and might not be comparable with other studies, which were focused mainly on undergraduate students.

Further analysis of the results of this study showed that there was moderate and significant correlation present between assessment by examiners and candidates, whereas the correlation between examiners and SPs and SPs and candidates was very low and not significant, which again demonstrates that there is a difference in opinion between examiners and SPs regarding the level of performance of candidates. The results by self-assessment and examiners assessment in our study are similar to another study's results (Jahan et al., 2011) on undergraduates. The results of the two studies however cannot be truly compared, as our study was conducted on experienced general practitioners.

Conclusion

Despite its limitations due to a relatively small sample size and small number of stations, with limited training of SPs as assessors, this study has highlighted an important issue, that the assessment of communication skills and empathy in an OSCE by examiners may not be reliable and could be different from SPs' opinion. This highlights the need for developing a system to involve simulated patients in the assessment process. Further research is needed on a much larger sample size and greater number of stations, to evaluate, whether SPs should be involved actively in the whole process of assessment in terms of reliability of communication skills assessment, time management and cost-effectiveness.

REFERENCES

- Allen, R., Heard, J., & Savidge, M. (1998). Global ratings versus checklist scoring in an OSCE. *Academic Medicine*, *73*, 597-598. doi:10.1097/00001888-199805000-00067
- Bianchi, F., Stobbe, K., & Eva, K. (2008). Comparing academic performance of medical students in distributed learning sites: The McMaster experience. *Medical Teacher*, *30*, 67-71. doi:10.1080/01421590701754144
- Bokken, L., Linssen, T., Scherpbier, A., Van der Vleuten, C., & Re-thans, J. J. (2009). Feedback by simulated patients in undergraduate medical education: A systematic review of the literature. *Medical Education*, *43*, 202-210. doi:10.1111/j.1365-2923.2008.03268.x
- Bokken, L., van Dalen, J., & Re-thans, J. J. (2010). The case of "Miss Jacobs": Adolescent simulated patients and the quality of their role playing, feedback, and personal impact. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, *5*, 315-319.
- Brannick, M. T., Erol-Korkmaz, H. T., & Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Medical Education*, *45*, 1181-1189. doi:10.1111/j.1365-2923.2011.04075.x
- Chumley, H. S. (2008). What does an OSCE checklist measure? *Family Medicine*, *40*, 589-591.
- Fischbeck, S., Mauch, M., Leschnik, E., Beutel, M. E., & Laubach, W. (2011). Assessment of communication skills with an OSCE among first year medical students. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *61*, 465-471. doi:10.1055/s-0031-1291277
- Harden, R. M., & Gleeson, F. A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education*, *13*, 41-54. doi:10.1111/j.1365-2923.1979.tb00918.x
- Hatala, R., Marr, S., Cuncic, C., & Bacchus, C. M. (2011). Modification of an OSCE format to enhance patient continuity in a high-stakes assessment of clinical performance. *BMC Medical Education*, *11*, 23-28. doi:10.1186/1472-6920-11-23
- Jahan, F., Sadaf, S., Bhanji, S., Naeem, N., & Qureshi, R. (2011). Clinical skills assessment: Comparison of student and examiner assessment in an objective structured clinical examination. *Education for Health (Abingdon, England)*, *24*, 421.
- Mazor, K. M., Ockene, J. K., Rogers, H. J., Carlin, M. M., & Quirk, M. E. (2005). The relationship between checklist scores on a communication OSCE and analogue patients' perceptions of communication. *Advances in Health Sciences Education*, *10*, 37-51. doi:10.1007/s10459-004-1790-2
- McNaughton, N., Tiberius, R., & Hodges, B. (1999). Effects of portraying psychologically and emotionally complex standardized patient roles. *Teaching and Learning in Medicine*, *11*, 135-141. doi:10.1207/S15328015TL110303
- Moineau, G., Power, B., Pion, A. M., Wood, T. J., & Humphrey-Murto, S. (2011). Comparison of student examiner to faculty examiner scoring and feedback in an OSCE. *Medical Education*, *45*, 183-191. doi:10.1111/j.1365-2923.2010.03800.x
- Pierre R. B., Wierenga A., Barton M., Thame K., Branday J. M., & Christie C. D. C. (2005). Student self-assessment in a paediatric objective structured clinical examination. *West Indian Medical Journal*, *54*, 144-148.
- Regehr G., & Eva K. (2006). Self-assessment, self-direction and self-regulating professional. *Clinical Orthopedics Related Research*, *449*, 34-48.
- Regehr, G., Freeman, R., Robb, A., Missiha, N., & Heisey, R. (1999). OSCE performance evaluations made by standardized patients: Comparing checklist and global rating scores. *Academic Medicine*, *74*, 135-137. doi:10.1097/00001888-199910000-00064
- Robb K. V., & Rothman A. (1985). The assessment of history-taking and physical examination skills in general internal medicine residents using a checklist. *Royal College of Physicians and Surgeons of Canada*, *20*, 45-48.
- Rosen, K. R. (2008). The history of medical simulation. *Journal of Critical Care*, *23*, 157-166. doi:10.1016/j.jcrr.2007.12.004
- Schwartzman, E., Hsu, D. I., Law, A. V., & Chung, E. P. (2011a). Assessment of patient communication skills during OSCE: Examin-

- ing effectiveness of a training program in minimizing inter-grader variability. *Patient Education and Counseling*, 83, 472-477. [doi:10.1016/j.pec.2011.04.001](https://doi.org/10.1016/j.pec.2011.04.001)
- Sijaja, M., Romi, D., & Prka, Z. (2006). Medical students' clinical skills do not match their teachers' expectations: Survey at Zagreb University School of Medicine, Croatia. *Croatian Medical Journal*, 47, 169-175.
- Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment Evaluation in Higher Education*, 27, 501-510. [doi:10.1080/0260293022000020273](https://doi.org/10.1080/0260293022000020273)
- Thistlethwaite, J. E. (2002). Developing an OSCE station to assess the ability of medical students to share information and decisions with patients: Issues relating to interrater reliability and the use of simulated patients. *Education for Health*, 15, 170-179. [doi:10.1080/13576280210133099](https://doi.org/10.1080/13576280210133099)
- Townsend, A. H., McIlvenny, S., Miller, C. J., & Dunn, E. V. (2001). The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Medical Education*, 35, 841-846. [doi:10.1046/j.1365-2923.2001.00957.x](https://doi.org/10.1046/j.1365-2923.2001.00957.x)
- Watson, M. C., Skelton, J. R., & Bond, C. M. (2004). Simulated patients in the community pharmacy setting. *Pharmacy World & Science*, 26, 32-37. [doi:10.1023/B:PHAR.0000013467.61875.ce](https://doi.org/10.1023/B:PHAR.0000013467.61875.ce)