

Concurrences and Differences between Faculty Staff and Standardized Patients in the Assessment of Medical Students in the Post-Clinical Clerkship Objective Structured Clinical Examination

AYAKO MAKUUCHI, YASUHIKO TAKEMOTO, IKUJI SHIMAZAKI,
HIROKI NAMIKAWA, MASANORI KOBAYASHI, SHIGEKI KINUHATA,
NORIKO KAMATA, HIROMITSU TOYODA, YOSHIHIRO TOCHINO,
MINA MORIMURA, and TAICHI SHUTO

Citation	Osaka City Medical Journal.
Issue Date	2018-06
Type	Journal Article
Textversion	Publisher
Right	© Osaka City Medical Association. https://osakashi-igakukai.com/ .

Placed on: Osaka City University Repository

Concurrences and Differences between Faculty Staff and Standardized Patients in the Assessment of Medical Students in the Post-Clinical Clerkship Objective Structured Clinical Examination

AYAKO MAKUUCHI, YASUHIKO TAKEMOTO, IKUJI SHIMAZAKI, HIROKI NAMIKAWA,
MASANORI KOBAYASHI, SHIGEKI KINUHATA, NORIKO KAMATA, HIROMITSU TOYODA,
YOSHIHIRO TOCHINO, MINA MORIMURA, and TAICHI SHUTO

*Department of Medical Education and General Practice,
Osaka City University Graduate School of Medicine*

Abstract

Background

Variations in assessments of medical students by faculty staff and standardized patients (SPs) have not been clarified by previous studies. We examined the concurrences and differences between and among these evaluators for medical students sitting the Post-Clinical Clerkship Objective Structured Clinical Examination (Post-CC OSCE).

Methods

The participants were 94 sixth-year Osaka City University Medical School students who were taking the Post-CC OSCE. They were each examined by three faculty staff and one or two SPs using a global assessment and six classified assessments. We analysed the average global/classified assessment scores for all 94 students, the average classified assessment scores for the students for whom differences exceeding 2 points between the average global assessment scores by the faculty staff and those by the SPs were not found, and those for the students for whom such differences were found using the Mann-Whitney U test.

Results

No significant differences were found regarding the average global and classified assessment scores among the faculty staff and among SPs. However, for six students, a difference exceeding 2 points was found between the global assessments by the faculty staff and those by the SPs. Five of these six students received ratings from faculty staff that were more than 2 points higher than those given by the SPs.

Conclusions

The SPs might have provided multilateral assessments from different standpoints than the faculty

Received January 31, 2017; accepted November 28, 2017.

Correspondence to: Yasuhiko Takemoto, MD, PhD.

Department of Medical Education and General Practice, Osaka City University Graduate School of Medicine,
1-4-3 Asahimachi, Abeno-ku, Osaka 545-8585, Japan

Tel: +81-6-6645-3797; Fax: +81-6-6645-3796

E-mail: yatakemoto@med.osaka-cu.ac.jp

staff for a minority of the participating students. Further investigations are warranted to establish the criteria of adequate assessments of clinical competencies using the Post-CC OSCE.

Key Words: Faculty; Standardized patients; Assessment; Post-Clinical Clerkship
Objective Structured Clinical Examination

Introduction

The objective structured clinical examination (OSCE) was introduced in the 1970s to solve existing problems regarding the assessment of clinical competencies in medical students^{1,2}. The OSCE has since been introduced all over the world and is recognized as a feasible approach to medical student assessment during the different phases of education, including the early and later years of the undergraduate program, as well as before clinical clerkship (CC) and after (Post-CC).

Medical training has traditionally depended on patient contact. However, changes in the healthcare system coupled with concerns about the lack of objectivity or standardization of clinical examinations has led to the introduction of simulated or standardized patients (SPs), who are usually lay people trained to portray patients with specific conditions in a realistic and sometimes standardized way^{2,3}. SPs are used for assessment of clinical competencies, for instance, in the OSCE or in the Post-CC OSCE.

Although the OSCE before CC for third- or fourth-year medical students was officially introduced in 2005 in Japan by the Common Achievement Tests Organization, the common Post-CC OSCE was not officially instituted. Therefore, the Post-CC OSCE is currently being administered at nearly half of all medical schools in Japan as a trial examination. Each school determines the design, content, and assessment of the exam individually⁴⁻⁹.

While several reliable formative assessment methods such as the 360-degree assessment, an assessment enforced multilaterally by multiple evaluators, have been proposed and developed for use in medical education¹⁰, generally only faculty staff evaluate medical students at their own medical schools. Consequently, few studies have examined student evaluations by SPs¹¹. In addition, no previous studies have reported on how both faculty staff and SPs simultaneously evaluate medical students' Post-CC OSCE performances. It is inferred that there will be some differences in the evaluation between faculty staff and SPs if medical students are evaluated multilaterally. The purpose of this study is to clarify the concurrences and differences in the assessment of medical students among faculty staff, among SPs, or between faculty staff and SPs.

Methods

Participants

A total of 94 sixth-year students of Osaka City University Medical School taking the Post-CC OSCE participated in this study, which consisted of the medical interview of an SP, taking an SP's pulse rate, and measuring the blood pressure of a simulator. The students were divided into eight groups: six groups of 12 students (booths 1 to 6) and two groups of 11 students (booths 7 and 8).

The study protocol adhered to the Declaration of Helsinki. Informed consent was obtained from all 94 students prior to participation in the study through an opt-out methodology that was approved by the ethics committee of Osaka City University.

Examinations

Each group was allocated to a booth, where the students took the three aforementioned

examinations. The students were guided to their booth one at a time, one minute prior to the start of the examination, where they read the instructions (Fig. 1). One minute later, the student began the interview with an SP. After completing the interview, the student measured the SP's pulse rate and the blood pressure of a simulator. The examiners consisted of three faculty staff from Osaka City University Medical School and one or two SPs per booth. Fifteen SPs participated in the examination. Then, two SPs were allocated to booths 1 to 7 and one SP was allocated to booth 8. They evaluated the students using an assessment sheet that consisted of a global assessment and six classified assessments in a four-level system (Fig. 2).

Statistical analysis

We replaced A (excellent) with 3 points, B (fair) with 2 points, C (good) with 1 point, and D (poor) with 0 points for the statistical analysis of the global and classified assessments. Then, we calculated the average global and classified assessment scores given by the faculty staff or SPs separately, and analysed the results for all 94 students using the Mann-Whitney U test.

Although the students were evaluated using a four-level system of A, B, C, and D, none of the students were evaluated as a D; thus, the students were essentially evaluated using a three-level system in the global assessment. Therefore, we defined the differences exceeding 2 points as the largest differences among the average global assessment scores by the faculty staff, among those by the SPs, and between the faculty staff and SPs. Then, we searched for student scores featuring the differences exceeding 2 points. When we searched for the differences among the average global assessment scores by the SPs, we excluded the students who were examined in booth 8 with one SP.

【Setting & your task】

- A 60-year-old patient whose name is Taro Yamamoto (male) or Hanako Yamamoto (female) presents to our hospital and is waiting for you outside of your booth.
 - You have one minute to read the instructions.
 - After the minute is up, guide the patient to have a seat in front of you in your booth.
 - 1) Interview the patient.
 - 2) Take the patient's pulse rate.
 - 3) Measure the blood pressure of the simulator in front of you.
- Perform all three procedures within 10 minutes.

Figure 1. Instruction sheet for the examination.

① Global assessment (A, excellent; B, fair; C, good; D, poor)

A	B	C	D
---	---	---	---

② Classified assessments (A, excellent; B, fair; C, good; D, poor)

	A	B	C	D
Keeps a clean appearance	A	B	C	D
Leads patients to take a seat and has a natural conversation	A	B	C	D
Listens to patients' stories carefully	A	B	C	D
Polite and clear-cut phrasing	A	B	C	D
Maintains conversational eye contact	A	B	C	D
Nods and responds to patients naturally	A	B	C	D

Figure 2. An assessment sheet consisting of a global assessment and six classified assessments in a four-level system.

There were two types of differences between the average global assessment scores by the faculty staff and those by the SPs: the differences exceeding 2 points between the highest assessment scores by the faculty staff and the lowest assessment scores by the SPs, or the differences exceeding 2 points between the lowest assessment scores by the faculty staff and the highest assessment scores by the SPs.

Subsequently, we analysed the average classified assessment scores using the Mann-Whitney U test for the students for whom differences exceeding 2 points between the average global assessment scores by the faculty staff and those by the SPs were not found and for the students for whom such differences were found.

Results

No significant differences were found in the average global and classified assessment scores among the faculty staff and among the SPs for all 94 students (Table 1). While differences exceeding 2 points in the average global assessment scores were not present among the faculty staff or the SPs, differences exceeding 2 points between the average global assessment scores by the faculty staff and those by the SPs were found for six out of 94 students. The faculty staff rated five out of six students more than 2 points higher than the SPs did. Therefore, for 88 out of 94 students, differences exceeding 2 points between the average global assessment scores by the faculty staff and those by the SPs were not found. For these 88 students' examinations, only the average classified assessment scores for "Keeps a clean appearance" were significantly different between the faculty staff and the SPs (Table 2, section 1). We found five students with differences exceeding 2 points between the

Table 1. Results of the Mann-Whitney U test for all 94 students

A=3, B=2, C=1, D=0		Evaluator	Effective number	Difference compared to actual number	Sum	Average	Standard deviation	p value
#Global assessment		Faculty staff	282	0	639	2.27	0.53	0.902
		SPs	177	0	402	2.27	0.54	
#Classified assessment	● Keeps a clean appearance	Faculty staff	282	0	644	2.41	0.56	0.475
		SPs	176	-1 ^a	436	2.38	0.56	
	● Leads patients to take a seat and has a natural conversation	Faculty staff	282	0	661	2.34	0.55	0.072
		SPs	176	-1 ^a	397	2.26	0.59	
	● Listens to patients' stories carefully	Faculty staff	281	-1 ^a	686	2.44	0.58	0.922
		SPs	176	-1 ^a	419	2.38	0.58	
	● Polite and clear-cut phrasing	Faculty staff	282	0	693	2.46	0.52	0.055
		SPs	176	-1 ^a	428	2.43	0.55	
	● Maintains conversational eye contact	Faculty staff	282	0	688	2.44	0.54	0.331
		SPs	176	-1 ^a	414	2.35	0.57	
	● Nods and responds to patients naturally	Faculty staff	282	0	678.5	2.41	0.56	0.475
		SPs	176	-1 ^a	418	2.38	0.56	

^a There was no check mark for one student. SPs, Standardized Patients.

Table 2. Results of the Mann-Whitney U test for the students for whom differences exceeding 2 points were not found and for whom such differences were found

1) The 88 students for whom differences exceeding 2 points were not found							
A=3, B=2, C=1, D=0	Evaluator	Effective number	Difference compared to actual number	Sum	Average	Standard deviation	p value
● Keeps a clean appearance	Faculty staff	264	0	603	2.28	0.56	0.000
	SPs	166	-1 ^a	415	2.50	0.53	
● Leads patients to take a seat and has a natural conversation	Faculty staff	264	0	620	2.35	0.56	0.203
	SPs	166	-1 ^a	377	2.27	0.59	
● Listens to patients' stories carefully	Faculty staff	263	-1 ^a	641	2.44	0.58	0.441
	SPs	166	-1 ^a	400	2.41	0.55	
● Polite and clear-cut phrasing	Faculty staff	264	0	650	2.46	0.52	0.843
	SPs	166	-1 ^a	407	2.45	0.52	
● Maintains conversational eye contact	Faculty staff	264	0	648	2.45	0.54	0.191
	SPs	166	-1 ^a	396	2.39	0.54	
● Nods and responds to patients naturally	Faculty staff	264	0	638.5 ^b	2.42	0.55	0.681
	SPs	166	-1 ^a	398	2.40	0.55	
2) The 5 students for whom differences exceeding 2 points were found							
A=3, B=2, C=1, D=0	Evaluator	Effective number	Difference compared to actual number	Sum	Average	Standard deviation	p value
● Keeps a clean appearance	Faculty staff	15	0	36	2.40	0.51	0.042
	SPs	8	0	16	2.00	0.00	
● Leads patients to take a seat and has a natural conversation	Faculty staff	15	0	34	2.27	0.46	0.023
	SPs	8	0	14	1.75	0.46	
● Listens to patients' stories carefully	Faculty staff	15	0	40	2.67	0.49	0.006
	SPs	8	0	13	1.63	0.74	
● Polite and clear-cut phrasing	Faculty staff	15	0	37	2.47	0.52	0.077
	SPs	8	0	15	1.88	0.83	
● Maintains conversational eye contact	Faculty staff	15	0	35	2.33	0.49	0.003
	SPs	8	0	12	1.50	0.53	
● Nods and responds to patients naturally	Faculty staff	15	0	36	2.40	0.51	0.011
	SPs	8	0	14	1.75	0.46	

^a There was no check mark for one student. ^b One student received both an A and a B, so we regarded their score as 2.5. SPs, Standardized Patients.

highest global assessment scores by the faculty staff and the lowest global assessment scores by the SPs and one student with a difference exceeding 2 points between the lowest scores by the faculty staff and the highest scores by the SPs. For these five students, five out of six average classified

assessment scores significantly differed between the faculty staff and the SPs (Table 2, section 2).

Discussion

In the present study, we examined the concurrences and differences in the assessment of medical students among faculty staff, among SPs, or between faculty staff and SPs.

Assessments among the faculty staff and among the SPs

For all 94 students, there were no significant differences in the average global and classified assessment scores among the faculty staff and among the SPs. This finding may indicate that the validity among the assessments by the faculty staff and among those by the SPs is ensured.

Assessments between the faculty staff and the SPs

We found differences exceeding 2 points between the average global assessment scores by the faculty staff and those by the SPs for the six out of 94 students although we did not find such differences for remaining 88 students. The six students were evaluated at different booths, not at one particular booth. This suggests that the differences in the average global assessment scores for these six students were unlikely to be biased by the difference of the booths.

Differences were not found exceeding 2 points between the average global assessment scores by the faculty staff and those by the SPs in 88 out of 94 students. One interpretation of the finding is that the simultaneous evaluations by both the faculty staff and the SPs might have not been necessary, considering the concurrences between the assessments by the faculty staff and those by the SPs for most of the participating students. Another interpretation is that the SPs might have provided multilateral assessments from different standpoints than the faculty staff or supported the faculty staff's assessments, considering the differences between the assessments by the faculty staff and those by the SPs for a minority of the participating students, although the differences were not statistically significant.

For the five students who were rated more than 2 points higher by the faculty staff than by the SPs in the average global assessment, five average classified assessment scores significantly differed between the faculty staff and the SPs. This finding suggests that the global assessments by the faculty staff and the SPs may be closely associated with each group's classified assessments.

In the six classified assessments of those five students, 'Polite and clear-cut phrasing' was the only item that did not show a significant difference between the average scores by the faculty staff and those by the SPs, with the remaining five classified assessments being rated higher by the former than the latter. The three out of five classified assessments, 'Listens to patients' stories carefully', 'Nods and responds to patients naturally', and 'Maintains conversational eye contact', that tended to be subjectively evaluated showed larger differences between the assessments by the faculty staff and those by the SPs than the remaining two classified assessments. Meanwhile, the remaining two assessments, 'Keeps a clean appearance' and 'Leads patients to take a seat and has a natural conversation', that tended to be objectively evaluated showed smaller differences between them than the three assessments described above. Hence, the differences exceeding 2 points between the global assessments by the faculty staff and those by the SPs may be mainly attributed to the three classified assessments that tended to be subjectively evaluated.

Whelan et al reported that real patients or persons trained to be patients (such as the SPs who participated in this study) were the most appropriate evaluators of doctor-patient communications, and that clinical experts (such as the faculty staff who participated in this study) were the most

appropriate assessors of clinical problem-solving skills¹²⁾. Therefore, the differences exceeding 2 points between the average global assessment scores by the faculty staff and those by the SPs may be attributed to different levels of pertinence in the evaluation of doctor-patient communications.

In this study, the faculty staff and SPs simultaneously evaluated the same students. While there were some disadvantages to this approach, such as the burdens of the SPs themselves in the examination or the training of the SPs prior to the examination, there might have been advantages such as multilateral evaluations by multiple evaluators in which the SPs, who have been described as more appropriate evaluators of doctor-patient communications, as well as the faculty staff, were engaged in the examination.

Limitations

First, the subjects of the present study were limited to sixth-year students from our medical school. Further examinations of students from other medical schools or grades are warranted.

Second, the SPs who participated in the Post-CC OSCE in this study had been trained for several years and already had the opportunity to evaluate the medical interviewing skills of students prior to their participation in the exam. However, it was the first time that almost all faculty staff had engaged in the Post-CC OSCE and rated the medical interviewing skills of students. Therefore, the difference in proficiency levels between the faculty staff and the SPs may be linked to the difference in evaluating the students.

Third, the clinical competencies of the students should be similar, to compare the assessments by the faculty staff with those of the SPs accurately, although this is extremely difficult due to the diversity of the students.

In conclusion, no significant differences were found between the average global and classified assessment scores by the faculty staff and those by the SPs for all 94 students. However, differences of more than 2 points were found between the global assessments by the faculty staff and those by the SPs for six students. The SPs might have provided multilateral assessments from different standpoints than the faculty staff for a minority of the participating students. Further investigations are warranted to establish adequate assessments of clinical competencies in the Post-CC OSCE.

References

1. Davis MH. OSCE: the Dundee experience. *Med Teach* 2003;25:255-261.
2. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Med Teach* 2003;25:262-270.
3. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42. *Med Teach* 2009;31:477-486.
4. Yoshida M. Implementation and challenges of Objective Structure Clinical Examination after clinical clerkship. *Igaku Kyoiku/Medical Education (Japan)* 2015;46:18-22. (In Japanese)
5. Yukami S, Kanazawa N, Hosoi A, Morita H, Marutani R, Ogawa M, et al. Experiences at the advanced OSCE Osaka trial. *Igaku Kyoiku/Medical Education (Japan)* 2005;36:113-118. (In Japanese)
6. Imanishi H. A brief overview of the Objective Structured Clinical Examination for medical students. *Acta Med Hyogo* 2014;38:51-55. (In Japanese)
7. Deguchi H, Hayashi T, Terasaki F, Ukimura A, Kitaura Y, Tsuda T, et al. Assessment of a system for evaluating clinical skills in cardiology with the Objective Structured Clinical Examination at the end of bedside learning. *Igaku Kyoiku/Medical Education (Japan)* 2004;35:245-253. (In Japanese)
8. Miyamoto M, Sugino S, Shimamoto F, Deguchi H, Kitaura Y, Yoneda H. Study of correlations between the examinees' results on the basic objective structured clinical examination (OSCE), computer-based testing, and the advanced OSCE. *Igaku Kyoiku/Medical Education (Japan)* 2007;38:399-405. (In Japanese)
9. Suzuki K, Naruse H. Advanced Objective Structured Clinical Examination trial at Hyogo College of Medicine. *Igaku Kyoiku/Medical Education (Japan)* 2007;38:103-110. (In Japanese)

10. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No 31. *Med Teach* 2007; 29:855-871.
11. Kondo T, Ichimura K, Ito K, Takahashi Y, Numaguchi C, Kuroda Y. The analysis of the difference between the evaluation given by different examiners on OSCE. *Acta Scientiarum Valetudinariae Universitatis Praefecturae Ibarakiensis* 2011;16:1-11. (In Japanese)
12. Whelan G, Boulet JR, McKinley DW, Norcini JJ, van Zanten M, Hambleton RK, et al. Scoring standardized patient examinations: lessons learned from the development and administration of the ECFMG Clinical Skills Assessment CSA®. *Med Teach* 2005;27:200-206.