

ORIGINAL ARTICLE

Volume 17 Issue 3 2025

DOI: 10.21315/eimj2025.17.3.7

ARTICLE INFO

Submitted: 16-05-2024

Accepted: 05-05-2025

Online: 30-09-2025

Exploring Factors that Influence Faculty and Simulated Patients' Assessment of Student's Communication Skills during Objective Structured Clinical Examination: A Mixed Methods Study

Ranila Ishani Sirisinghe¹, Pathiyil Ravi Shankar², Chew-Fei Sow^{1,2}, Liao Jia Li¹, Manimagalai Krishnan³, Ramanathan Subramanian¹, Chandrashekhar T Sreeramareddy⁴

¹Clinical Skills Department, School of Medicine, IMU University, Kuala Lumpur, MALAYSIA

²IMU Centre of Education (ICE), IMU University, Kuala Lumpur, MALAYSIA

³Clinical Skills and Simulation Centre, IMU University, Kuala Lumpur, MALAYSIA

⁴Public Health and Community Medicine, School of Medicine, IMU University, Kuala Lumpur, MALAYSIA

To cite this article: Sirisinghe RI, Shankar PV, Sow CF, Li LJ, Krishnan M, Subramanian R, et al. Exploring factors that influence faculty and simulated patients' assessment of student's communication skills during objective structured clinical examination: a mixed methods study. *Education in Medicine Journal*. 2025;17(3):75-92. <https://doi.org/10.21315/eimj2025.17.3.7>

To link to this article: <https://doi.org/10.21315/eimj2025.17.3.7>

ABSTRACT

Simulated patients (SPs) and faculty are integral assessors tasked with grading medical students' communication skills in objective structured clinical examinations (OSCEs). Understanding their perspectives on this evaluation process is vital for ensuring fairness and impartiality. We conducted a comprehensive analysis, investigating both the quantitative interrater agreement between faculty and trained SPs' grades on students' communication skills and qualitatively exploring the factors influencing their grading decisions. Anonymised communication skill grades from 376 students were obtained and analysed. There was a lack of agreement between SPs and faculty grades at different stations. Furthermore, we explored the factors that influence their grading practices through focused group discussions with SPs and faculty examiners. Shared views between SPs and faculty regarding factors influencing grading included the scale of examinations and students' semesters. Large-scale examinations cause mental fatigue among examiners, and students from higher semesters of study are graded more strictly. The factors that did not influence grading were the course fees, consequences of failing or demotivating students, and examination rules. SPs and faculty held differing views on the influence of their roles and student characteristics, notably grooming and appearance. These disparities expose students to diverse patient perspectives they will encounter in their future roles as healthcare professionals, highlighting the complexity of communication skills assessment in OSCEs and the necessity for a comprehensive approach.

Keywords: *Assessment, Communication skills, Simulated patients, Faculty, OSCE*

CORRESPONDING AUTHOR

Chew-Fei Sow, Clinical Skills Department, School of Medicine, IMU University, 126, Jln Jalil Perkasa 19, Bukit Jalil, 57000 Kuala Lumpur, Malaysia

Email: chewfei_sow@imu.edu.my

INTRODUCTION

Objective structured clinical examinations (OSCEs) are used to evaluate a wide range of clinical competencies, including clinical decision-making, communication, history-taking, physical examination, procedural skills, and professionalism (1). The OSCE is structured and standardised, which means that all the students are tested on the same tasks in the same manner. Both simulated patients (SPs) and faculty members are commonly used as assessors for communication skills in OSCE. Using SPs can help reduce the variability that would arise if real patients were used and improve the reliability and validity of assessment (2, 3). In this context, training SPs is critical to ensure that all SPs provide consistent and realistic evaluations across multiple encounters. The training of SPs improves OSCE quality and the overall satisfaction of both SPs and students (4).

Effective communication between patients and doctors is essential for better patient satisfaction and successful outcomes. However, patients and medical educators might have different expectations regarding good communication skills, empathy, and professionalism. The communication skills emphasised by faculty might differ from those emphasised by SPs (5). Hence, the use of SPs as assessors could provide additional information about students' performance and complement the grading of faculty. The incorporation of SP ratings in summative assessments may produce doctors with a holistic approach (6).

IMU University has included trained SPs' ratings of student communication skills in progression point OSCEs during semesters 3 and 5. Several studies have investigated the differences between SPs and faculty assessments of communication skills among medical students in OSCEs (7–10). SP's empathy grades were higher than those of the faculty when assessing medical students during their family medicine clerkship (7). This is also supported by another study that showed that SPs gave higher scores than did the obstetrics and gynaecology faculty when assessing postgraduate trainees' communication skills in a formative OSCE (10). A significant difference in assessments of the verbal communication of postgraduate general practitioners between SPs and examiners was noted (9). Physician examiners and SPs provide poorly correlated ratings of medical student interviewing skills (11). Both groups appeared to have different perspectives and perceptions of the students' interviews. However, few studies have explored how SPs/faculty grade students and what factors contribute to the discrepancies in communication skills assessment during OSCE between SPs and faculty.

With this background, this study was conducted to determine the interrater agreement between faculty and SP grades of communication skills in OSCE and to explore the factors they considered when grading the examinees.

METHODS

Setting

At IMU University, the medical preclinical curriculum consisted of progression point examinations in the third semester, as well as the fifth semester before medical students entered their clinical phase of medical school. The examinations include theory and clinical examinations (OSEs), in which SPs (only communication skills) and faculty assess student performance.

Since 2017, IMU University has implemented structured training for all SPs participating in teaching, learning and assessments. The programme consists of the following: basic training on the roles and responsibilities of an SP, training on giving feedback, case-based training, and evaluation of SP performance. Faculty members are also trained before becoming OSCE examiners. This includes an introduction to the OSCE, providing effective feedback, training on the online assessment system used at the university, and hands-on activities to calibrate assessments via the global rating of students.

In the OSCE, there are various stations, including history taking, physical examinations, and information provision. At each station, the assessment rubrics differ depending on the expected competencies and learning outcomes of the station.

Study Design

A mixed-methods approach was selected. First, the interrater agreement between faculty and trained SPs in grading student communication skills during OSCE was measured from examination records. We then conducted focus group discussions to explore the factors that SPs and faculty members considered when evaluating students' communication skills. An explanatory sequential mixed-methods design was used, starting with quantitative research to investigate whether any significant differences existed in scores between the SPs and faculty. A qualitative study was subsequently conducted to explore the factors that may have influenced the scoring, with the goal of understanding the reasons behind any differences identified.

Quantitative study

After written permission was obtained, grades given by the SPs and OSCE examiners were captured in the online assessment system from the examination unit at 14 OSCE stations. The data were obtained from two cohorts comprising 176 (semester 3) and 200 (semester 5) medical students. This selection was underpinned by three critical considerations: (a) both cohorts belonged to the old medical curriculum, which was in place until 2021, when IMU University introduced a revised curriculum; (b) to ensure a consistent educational background, the exams were conducted during the COVID-19 pandemic, providing a uniform period for evaluation; and (c) most importantly, the selection allowed for comparative analysis across two distinct progression points within the curriculum. For all the stations, the SPs marked each examinee on a scale of 0 to 2. As the rating scales provided for OSCE examiners and SPs differ, we converted the scores into a percentage score as a standard scale to calculate interrater agreement. For each OSCE station, the percentage agreement and Cohen's kappa coefficient were calculated as measures of interrater agreement via StataMP 11.0 (12). Of the 14 stations analysed, 10 had histories in which communication skills were scored in the range of 0 to 4, and 4 were physical examination stations scored from 0 to 3. The data displayed indicate that the score (mode) most frequently awarded by both the SP and the examiners is 2 points. The percentage agreement between faculty and SPs ranged from 0% to 62.5%, whereas Cohen's kappa coefficient ranged from -0.0053 to 0.198. This kappa statistic of 0 to 0.20 suggests no agreement between the two markers, as shown in Table 1.

Table 1: Description of the examination, content of OSCE stations, and interrater agreement between SPs and faculty

Cohort	Semester	Station	Station content	Agreement percentage	Cohen's Kappa	p-value
Cohort X	Semester 3 progression point examination	332	History taking and question	62.50	0.1529	0.0156
		333	Focused history and explanation	16.48	0.0555	0.0002
		334	History and question	12.50	0.0603	< 0.0001
		335	History	23.30	0.0372	0.0185
		337	Procedure information giving/communication	39.77	0.0969	< 0.0001
		339	History	11.93	0.0508	0.0002
Cohort Y	Semester 5 progression point examination	393	Physical examination	00.00	−0.0053	0.9230
		394	Physical examination	23.50	0.0853	< 0.0001
		395	History taking	24.50	0.0826	0.0002
		396	History taking	37.50	0.1979	< 0.0001
		397	Physical examination	26.50	0.1103	< 0.0001
		398	Physical examination	49.00	0.1167	< 0.0001
		399	History taking and information giving	36.00	0.1882	< 0.0001
		400	History taking and information giving	24.50	0.0720	0.0026

Qualitative study

Focus group discussions (FGDs) were conducted with SPs and faculty members. A total of eight FGDs were conducted (four each with SPs and faculty examiners). Both groups were divided into four groups on the basis of their years of experience as examiners (< 1 year, 1–5 years, 5–10 years and > 10 years). The initial phase established their definition of a good doctor, setting a threshold for their assessment standards. The factors that may have subsequently influenced their assessments during OSCE are discussed. The inclusion criteria included SP and faculty members who had undergone OSCE examiner training and had at least one prior experience as an assessor in the OSCE.

Data Collection and Tools

A study (13) on Canadian and Chinese high school teachers' grading decisions and practices shed light on the factors influencing assessment practices. Based on these findings, the guiding questions for the FGDs with SPs and faculty members assessing students' communication skills in OSCEs were constructed. The semi-structured interviews were guided by eight domains: (a) general questions on qualities of a good/bad doctor, personal communication experiences, and recommendations; (b) context and classroom management: subject variation in OSCE roles; (c) context and classroom management: grade-level considerations; (d) learning values: grades as motivation; (e) learning values: policy-based grading; (f) learning values: large-scale testing; (g) external pressures: administrators

and course fees; and (h) consequences of grade use and student factors. Example prompts included: "What qualities does a good doctor possess?" and "How do your roles as assessor affect your grading of students?" Two faculty members not involved in the study validated the guiding questions.

All eligible SPs and faculty members meeting the inclusion criteria were invited to participate. The first three authors facilitated the FGDs. The interviewers were medical doctors and educators in the health profession. They knew the participants as faculty colleagues and SPs working at the same university. Two interviewers were female, one was male, and they were experienced in qualitative research methods. The research was introduced to the participants, and a detailed study information sheet was provided. The participants were free to clarify any queries. A total of 16 faculty members participated, including one general surgeon, four basic scientists, one clinical skills nurse, eight clinical skills facilitators with medical qualifications, one ear, nose and throat (ENT) surgeon, and one emergency specialist. The faculty assesses the students during history-taking stations based on their verbal and nonverbal communication skills, fluency, clarity of articulation without medical jargon, and attentiveness to the patient while demonstrating genuine interest in their concerns. Faculty evaluate students' communication skills during physical examination stations on their ability to establish rapport, communicate clearly, and provide organised and attentive care, focusing on the overall quality of the examination and the student's approach to patient interaction. The 16 SPs included homemakers, other university students, estate planners, and retirees from various occupations. The SP evaluates the students' communication skills in history-taking stations based on their demonstration of politeness and respect, ability to create a comfortable environment during the interview/examination, understanding of the SP's questions and instructions, display of empathy, and responsiveness to the SP's concerns and worries. In assessing communication skills at physical examination stations, the SP evaluates the clarity of the student's instructions, the overall comfort level during the examination, and the student's ability to instill confidence in the SP throughout the interaction.

Written informed consent was obtained from all individuals. FGDs were conducted face-to-face or via Microsoft Teams and recorded with participants' consent. The face-to-face FGDs were conducted in a meeting room at the Clinical Skills Centre. Only the interviewer, a notetaker (one of the researchers) and the participants were present. The FGD guide served to guide the discussions, but the participants were allowed freedom to explore areas related to the topic. Each FGD lasted between 60 minutes and 70 minutes. Data collection was continued until saturation was achieved. Transcripts were generated via Microsoft Teams' transcription function and reviewed verbatim for accuracy and deidentification purposes, maintaining confidentiality and rigour throughout the process. The transcripts were read through at the end of each FGD to identify possible themes. Data saturation was achieved when no new themes were identified during the FGD. The transcripts were provided to the participants for member checking.

Data Analysis

Qualitative study

A thematic analysis approach was adopted for the qualitative study, as outlined by Braun and Clarke (14). We used a mixed-type (predominantly inductive) analysis. We grouped the points mentioned by the respondents according to the areas in the FGD guide. New themes

were created where needed. Each interview transcript was reviewed by two researchers, who identified key phrases and words to generate codes and themes for subsequent analysis. The finalisation of codes and themes occurred through discussions among all the researchers. Two of the interviewers were also examiners. We were careful to ensure that the personal views of these examiners did not influence the interview process. Each transcript was initially reviewed to gain a broad understanding of the data. Subsequently, the transcripts were meticulously examined line by line, focusing on both superficial meanings and implied content, with each segment summarised and labelled accordingly. Labels were formulated to encapsulate keywords and underlying concepts within the conversation. A comprehensive list of labels for each FGD was compiled, consolidating similar codes into categories. Transcript reviews were repeated independently and jointly by two researchers to ensure the accuracy and completeness of coding. Categories were then clustered to form themes, potentially addressing research questions. The first three authors carried out this process. All transcripts, codes, categories, and themes were then shared with all the researchers. All provided inputs on the categories and themes identified and agreed with the final themes mentioned. If needed, modifications were recommended and carried out until consensus was achieved.

RESULTS

For the qualitative part of the study, the themes that emerged were the qualities of a good doctor, the qualities of a bad doctor, personal experience with doctors as a patient, the influence of different roles on grading, whether the semester of the student impacts the grading, how student motivation affects grading, the effect of examination rules on grading, the impact of the scale and logistics of the OSCE on grading, the consequences of a failed grade on students, the discrepancy between SP and faculty scores, the influence of students’ personal characteristics on grades, the influence of student fees on grading, the influence of training on SP grading, leniency or otherwise grading, and other influences on grading. Specific factors were explored based on the study mentioned earlier (12), which examined the various factors that influence teacher grading. Table 2 shows the interview questions, the probes used and the themes identified.

Table 2: Interview questions, probes and themes identified

Interview question	Probes used	Characteristics mentioned/considered	Themes identified
What qualities does a good doctor possess?	Why do you say these are qualities of a good doctor?	Knowledgeable, competent, confident, attentive, care/caring	Qualities of a good doctor
	What is the basis for your choice/decision?		
What qualities does a bad doctor possess?	Why do you say these are qualities of a bad doctor?	Rotating doctors, long waiting time, do not have adequate patient, history, rude, not attentive, crossing professional boundaries.	Qualities of a bad doctor
	What is the basis for your choice/decision?		

(Continued on next page)

Table 2 (Continued)

Interview question	Probes used	Characteristics mentioned/considered	Themes identified
Have you ever had a good/bad experience communicating with a healthcare professional? If yes, explain.	Will you be comfortable sharing the situation? Why was the situation good/bad for you? How did the situation make you feel?	Language, compassion, tone of voice, eye contact. Prompt response (good experience); Lack of answers, not being taken seriously, inadequate information (bad experience)	1. Personal experiences with doctor 2. Qualities of a doctor
You are given different roles during the OSCE? Do these roles influence your grading of the examinees?	The roles played by a SP were enumerated if needed. The roles by a faculty (station developer, coordinator, student mentor)	SP made to feel comfortable by examinee, maintain focus, importance of exam for student	Influence of different roles on grading
Do the fact that students may be from different semesters (level) of study affect your grading of their communication skills in exams?	If needed, examples were provided. The fact that whether a semester five would be graded differently from a first semester student during the basic sciences and a final year and a third-year student in the clinical sciences was enquired.	Higher standards for senior semesters (fifth and final) Safe for patients	Influence of semester of students on grading
Are you concerned the marks you awarded to the students could motivate or demotivate the students?	If a student receives a failing grade, will it demotivate him or her? Do you consider this while grading them?	Careful choice of words, constructive comments, feedback sandwich, the scores will be known only later	1. Influence of examination grades on student motivation and its effect on grading 2. Influence of semester of students on grading
And how does this affect your grading of their performance?			
How does the OSCE SP/faculty rules affect your grading of their performance? If yes, which rules and how?	Examples of the rules were provided. The sequence of candidates (if a good candidate was followed by a weak one) and grading was explored	Maintain focus, commitment, realise importance for student	Influence of examination rules on grading
How does the scale, large or small, of OSCE (i.e., number of students, number of different SPs/examiners of the same station) affect your grading of their performance?	Scale, number of examinees, examiner fatigue on grading was explored	Maintain focus, commitment, realise importance for student	Influence of the scale of examination/logistics on grading

(Continued on next page)

Table 2 (Continued)

Interview question	Probes used	Characteristics mentioned/considered	Themes identified
How does knowing the course fees affect your grading?	Influence of high fees paid by students on grades?	No. Only think about the patient. Safe doctors	1. Influence of knowing the consequences of failure on grading
How does knowing the consequences of your grading at progression point affect your marking?	A fifth semester student if s/he does not pass cannot go to the clinics? A final year student cannot graduate. Does this affect your grading?		2. Influence of examination grades on student motivation
How about in the resit examinations?			3. Influence of semester of students on grading
How does the students' attire, accent and personal grooming affect your grading?	Specific examples like personal grooming, ethnic differences, accent were discussed	Clean and professional appearance: dress, hair, grooming	Influence of student characteristics on grading
Have you noticed discrepancies between the grades provided by faculty and SP examiners?	–	Yes. SP grades only small portion.	1. Discrepancy between faculty and SP grading
What do you do in this situation?		Difference in rubrics. Expectations.	2. Influence of student characteristics on grading
			3. Influence of examination grades on student motivation

Qualities of a Good Doctor

A good doctor should be compassionate, show empathy and have good bedside manners. S/he should be knowledgeable about medicine, listen carefully to the patient, be confident, and demonstrate professionalism. A faculty member mentioned that in Malaysia and other Asian countries, doctors do not properly introduce themselves to patients and that sometimes courtesy may be lacking. An SP mentioned,

(The doctor should have)...very good knowledge in his field... must also listen to the patient. In addition, also they have to have a good rapport with the patient and of course... Have to be polite and like one of the SPs said...you know, we have to be comfortable with them. (SP 128)

Qualities of a Bad Doctor

A bad doctor is unprofessional, not organised, impatient and does not listen to patients. They may cross professional boundaries and not treat their patients properly. Some respondents noted that doctors in government clinics and hospitals may not devote sufficient time to their patients, fail to maintain proper eye contact, and often change. However, their professional manners may not be appropriate. Long waiting times were also mentioned.

Personal Experiences with Doctors

The respondents had both good and bad experiences with doctors. They mentioned the reasons why they regarded a particular experience as good or bad. The relationship with other healthcare professionals and the hierarchy in the medical profession were also discussed. Language, compassion, tone of voice, eye contact, and empathy were considered. Prompt response and provision of correct information also resulted in a good experience. A respondent mentioned,

Therefore, I just rushed and in front of the emergency. I just left my car over there and threw the key to the valet parking people. OK, take it anywhere, I am, I have to take him into the emergency. The emergency doctor, she was literally, you know, indifferent. OK, fine. You can stand over here; we are doing our best. She was literally on phone. My dad was like dying of pain, and I knew that half a second was very important and crucial for his life. Half a second, it will make a difference. Please do something. Please push that button. Please push the code blue. Please do something. She was, Ohh yeah... she was so slow. I literally can burst(ed) out you know... what the hell? You are a doctor. (Faculty K)

Influence of Different Roles on Grading

SPs play different roles depending on the station. Some stations may be communication stations, whereas others may be procedures or physical examination stations. Male SPs are often used for physical examination stations. Having several students repeatedly examine the same SP can result in physical discomfort. Examples of abdominal examination and blood pressure measurement were mentioned. The way the student approached and performed at the station influences the grading.

One faculty member mentioned,

So I think that is very important in regard to assessing comm skills for us to be able to come away from our own roles and understand what is expected of the student in that particular station at that level. This is then used to calibrate and assess them accordingly at that level. Therefore, I think that our own roles should not truly influence too much. (Faculty S)

Influence of Cohorts/Semesters of Students on Grading

Generally, expectations are higher for fifth-semester students, as they will be entering the clinical phase, and for final-year students, who will be graduating. High standards were also expected from fifth-semester students appearing in the resit examinations. One SP mentioned, "After three years of education, you expect them to be more educated you know... So there...of course, the marking will be stricter...you do want doctors... good doctors.. (SP 3)

Influence of Examination Grade on Student Motivation and Its Effect on Grading

The respondents were not concerned about the influence of grade level on student motivation. They also provide constructive feedback on the grade and believe that this should help the students improve. The marks provided by the SP were lower than those provided by the faculty, and the respondents were more concerned about whether the feedback they provided would negatively influence the students. One faculty member disagreed and mentioned that he worried about the marking and the grades he provided.

In a way, they have to understand it is given to them for their future improvement. Even though it is low now, then they relate back with their feedback and what can be done to improve on and move the marks up. Therefore, I think that is more important than we feel, OK the students will be demotivated, so I should give a slightly higher mark, but later on, they're not learning out of that. (Faculty B)

Influence of Examination Rules on Grading

The examination rules and the rubrics were regarded as helping the grading process and improving the grading quality. The creation of a comfortable environment for the students was emphasised. The examiners examined the students sequentially at the OSCE station, and the influence of the first student on the grades of subsequent students was discussed.

Yes, we have our rubrics, but for me, I think it all starts with the first student that you assess because that will be your benchmark. So if... unfortunately the first student is one of the good students, then your subsequent marking will be based on that benchmark. So that is one challenge I face. So if your first student happens to be a poor student, then subsequently you have to go back and remark that, that the weaker students...because you have seen a weak student, all the rest are doing much better than him. So that benchmark is something that, umm, I find a challenge in maintaining. (Faculty V)

Influence of the Scale of Examination/Logistics on Grading

The large number of examinees and the long exam duration may influence grading. Examiner fatigue was mentioned, and examiners may feel tired towards the end of the day and after lunch. Maintaining the same concentration level was challenging, but the respondents were aware of the importance of the OSCE for the students and focused on providing fair and objective grading.

Even though I feel tired, you know, after the seven, eight students or ten students, I will just pop a mint and then I will drink some water, you know, and pull myself up again because I know that...It is far more important for them than for me, you know, so I have to be like my partner said on my toes. Yeah. So usually, I will pop a mint and drink some water. (SP34)

Influence of Knowledge of the Consequences of Failure on Grading

The respondents mentioned that this does not influence their grading. They mentioned that they may be stricter during formative exams and provide feedback for the students to improve. A few mentioned that they may be more stringent with students who resit the exam to ensure that they have the competence to be a doctor. A participant candidly explored the influence of personal relationships with students on grading.

I am not biased in any way, if at all. Maybe one of my mentees who has had problems or who I built a personal rapport with, and you know we have our favourites. We all have our favourites, whether you like it or not...Some students are good, and we automatically gravitate towards them. We want to encourage them. You have got weak students who actually reach out, and they are actually very...they actually want to improve, and you help them out. Therefore, what should we...be to build that personal relationship then? I think it will definitely affect you if you are grading them. Therefore, if I were in that situation, I would probably not grade that student. Ask somebody else to replace me for that. If I mean, I do not know. I have not come across it yet, but thus far, I do not have any personal favourites or students that I am particularly interested in and have a vested interest in them passing. (Faculty V)

Influence of Student Characteristics on Grading

Most respondents agreed that background, age, ethnicity, etc., do not affect grading, although there were mixed views on student grooming and attention. There may be some difficulty in understanding the speech of international students due to differences in accents and the terms used. The issue of students carrying out sensitive examinations on patients of the opposite gender and the comfort level of the students was mentioned. Obtaining a history such as a menstrual history or a sexual history may also be difficult for students in lower semesters. Whether the present generation of students has a different communication style because they spend so much of their time with electronic devices was also mentioned.

One of them was appearance, hair, the collar, the shirt. Some of them come very untidy.... So we nicely...we kindly tell them you know I think your...your hair should go be combed properly...because you are a doctor and you're going to be a doctor. Very important how you look. Yeah, you must look professional. (SP 12)

Influence of Student Fees on Grading

The respondents agreed that the fees paid by the students do not affect their grading. Students are graded according to their performance via a rubric, and economic considerations do not influence the process. The quality of the doctors produced and their competence are important characteristics.

The fees. I do not think so will affect because I look at the future. What are we producing from our side? The quality from our side because they're going to treat us at one point, maybe who knows, and we produce very poor group of doctors means at the end is going to impact care. (Faculty B)

Discrepancy between Faculty and SP Grading

There were differences between SP and faculty grades. SP grades only a total of two marks and focuses mainly on communication skills. They emphasised whether they were comfortable interacting with the student doctor as a patient. Faculty examiners account for a greater percentage of the total grade and focus mainly on the technical aspects of student performance.

We are standardised patients, so we already know our standard. In addition, then there is... the nurses, the coordinators and examiners know how we perform so, and sometimes we calibrate with the examiners...and this is...is it OK. Is it anything we want to change?

You know, it is already like what the examiners want us to do, truly. Therefore, in that moment, right. It's the students' skills just have to save themselves. In addition, you're grading on the basis of the skills there...sometimes, you know, the SPS are kind enough to actually...Help them eliminate certain things that they do not need to know, but yeah, sometimes the students do not get it. They do not hear it. Therefore, there's...there's so much we can do. (SP 61)

Some other findings from the FGDs are mentioned in Table 3.

Table 3: Other findings from the FGD

Theme	Quotes
Qualities of a good doctor	“Especially the doctors, need to have their effective communication skill which is crucial to explain the complex medical terminologies and...and to listen attentively this number (is) 1... number 2 empathy and compassion and number 3 you can say professionalism and last, I can say that the skills and the competence.” (Faculty I) “Good communication skill, demonstrate empathy. Be honest. Show evidence of altruism. Be punctual, especially if he or she is gonna practice uh, you know, show respect towards others.” (Faculty P)
Qualities of a bad doctor	“When you say bad experience is not truly bad experience then, but I think doctors should be careful with their choice of words. I mean, I mean it is not truly bad experience, but you know like, for example when my daughter was born, she was a bit small size, you know. Then, we take her good for medical check-up, then we ask the doctor... I mean naturally as first parents... as first time parents. Doctor, how is she? Oh, but she's OK now, but sixteen years later, you have to be...16 years later, you will be worried...We were so stunned...Doctor what you mean by 16 years later... you know what it oh, when she go party then you have to worry.” (SP65) “We just ask the doctor our first questions, but they get agitated or irritated with our questions.” (SP 66)
Personal experience with doctors	“She...The nurse brought in a file and that file was wrong and she threw it back at the nurse so that acting was a very uh, unethical. Therefore, after the incident, I did not...I changed my doctor's because she did not give much attention to my issues.” (SP 3) “He was very approachable very...with a smiling face. Very, very carefree and very softened up. The situation. Then, at last I...I choose to do the operation with him and after that everything goes well. So I feel connected ...because I feel connected with him and I felt comfortable.” (SP 34)
Influence of different roles of faculty and SP on grading	“So it is not so much of the role that I carry, but I actually zoom into his performance, so to speak...The way he handled the patient.” (SP34) “More to the nonverbal communication skills. The way they touch the patient, the way they handle the patient. Uh, do those things where rapport, empathy would be a bit more easier to assess as compared to the history taking stations where it is more about listening, questioning, the nonverbal kind of communication.” (Faculty Y)
Influence of cohort/semester of students on grading	“I think generally the expectation to be high because they have exposed to so many classes. They have also had hospital, you know, attachments with real patients, interactions with real patients and that is why my expectation I think is more and it is influenced by what I know how our standard of our teachings and that has been guided by the expectations from the clinical schools like partner medical schools.” (Faculty M)
Influence of examination rules on grading	“So sometimes it is very hard to give them marks because it totally blows the thing away when you keep on looking at the clock...So, I do not know whether it is a good idea to allow them to bring the clock.” (SP10)

(Continued on next page)

Table 3 (Continued)

Theme	Quotes
Influence of examination grades on student motivation, and its effect on grading	"I think actually yes, I'm concerned, but actually I have a personal concern. I do not like to fail anybody, but unfortunately some of them just do not do what they're supposed to do or they do not say what they're supposed to say. So, you just cannot give them their marks." (Faculty V) "(I do not give a zero) unless the student is truly, truly bad...I cannot hear them, you know? Totally 0 eye contact...you know...Even if I say, could you speak louder? And yet the student....did not improve, you know, I mean, I do not want a future doctor to be like that." (SP1)
Influence of the scale of examination/ logistics on grading	"Student number might influence due to examiner exhaustion at the end of the day, especially after the lunch ones. The focus that might be an issue...I do not think. I do not want to say that everything well and good. Sure. Therefore, when it goes with the long hours after sometime, surely the focus is bit off. Therefore, I have to truly truly recall and see and try to keep myself properly." (Faculty B) "You know, it is very (demanding) doing the palpation and all that on the stomach area. They did the light and the deep. That is where. So we all have to (be) clear. We do not eat any good food...right food...All this...we do not need to disturb the flow of the exam." (SP15)
Influence of knowing consequences of failure on grading	"We are looking at producing quality students and quality doctors, so it is not. That we should not be affected by all of these other things. Yep, so that is me." (Faculty S) "No, I think we must understand that we are training doctors, so dealing with patients' lives, right. Therefore, you, cannot weigh the cost of the course of or whether we are a private or a public university." (Faculty P)
Influence of student characteristics on grading	"I think you have to factor one thing in, when we are grading our students during OSCE and that is and their background, some of them are foreigners. They have an accent when they speak, our SP's can't understand their accent. Even sometimes I find it difficult to understand what they are...what they're saying, so I think you have to give them some leeway on that...And the choice of words also differs, from their background, so that is another thing I think it is not fair to mark them down because of these cultural differences. So it is difficult to remain neutral sometimes." (Faculty V) "The social interaction of the current generation is different to what we used to have. OK, now even in our homes, umm our children, when they're faced with strangers, they do not communicate at all. There are no pleasantries exchanged. Like what we used to do... right? They're just, they're just quiet. So now we have got them in a medical situation, in a medical curriculum and we are now asking them to come in, meet a stranger, you know, and then display that empathy, the pleasantries and so on." (Faculty G)
Influence of knowing student fees on grading	"I think as a SP, we have to grade according to the performance of the student. It is what I feel... not according to the fees." (SP 60)
Discrepancy between faculty and SP grading	"You should take a judgement call based on those things. I mean that that will introduce bias in your marking, right? Yeah. Therefore, I think that is not fair. I mean, yeah, you are assessing the students. Whether the SP gives a low mark, a high mark, does not matter. I mean, your decision as a marker is the one that is your job to be done last. Therefore, I think all these are all, uh, secondary things. It may play on your mind, but it should not affect your marking at any cost. I do not think that makes the assessment fair to the student." (Faculty P)

DISCUSSION

We found no agreement between the scores given by faculty and SPs. This contradicts the findings from one study that showed a positive correlation between SP and faculty grading of communication skills, with an almost near-perfect agreement (15). This study was also carried out with medical students during the OSCE; however, the students assessed were in their first and second years, whereas in our study, they were second- and third-year students. Another study reported substantial agreement between SP and faculty evaluations of communication skills among first-year pharmacy students completing four OSCE stations (16). Our findings align with the findings of multiple other studies that reported discrepancies between the ratings of SPs and faculty members. One study reported that while the interrater correlation between SPs and clinical tutors on empathy ratings was reasonably good, there were some differences in their ratings (17). Another study also revealed significant differences between SP and faculty scoring of dental students' communication skills (18). Although this study included different demographic data of the students, the difference is still significant, considering the similarities in training between dental and medical students. Other researchers have investigated the differences between SPs and physician examiner grades and noted discrepancies. However, further exploration of the factors contributing to this discrepancy has not been conducted (9, 11).

These grading variations reflect the differing approaches that SPs and faculty use in assessing communication skills. While SPs evaluate elements of the patient experience, including empathy, comfort, and how well the student acknowledges patient concerns, faculty often stress clinical communication competencies, such as structured questioning, clear instructions, and appropriate medical terminology. We believe that these methods work hand in hand, and students recognise the importance of offering support, demonstrating empathy, and taking into account patient concerns and expectations while also honing their clinical skills. This synergy enhances students' capacity to appreciate various perspectives and viewpoints. Grasping how assessment criteria differ and their influence on student feedback relies on recognising these distinct yet complementary outlooks. This understanding may reveal chances to improve alignment in assessment practices, thus ensuring that both patient-centred communication and clinical effectiveness are adequately addressed.

Simulated patients evaluate patient-centred communication, whereas faculty focus on clinical competencies. Although these viewpoints diverge, showing consensus would indicate that proficiency in one area bolsters competence in the other, underscoring the necessity for integrated communication training. The observed lack of consensus points to potential shortcomings, suggesting that students might excel in one area while facing challenges in another. Hodges et al. (19) assert that multisource assessments increase validity, but determining agreement is crucial to understanding whether these skills evolve concurrently. As no consensus has been reached, this highlights the requirement for targeted training to assist students in merging technical skills with empathetic, patient-centred care.

This study investigated the factors contributing to the SP and faculty grading of student communication skills in OSCE. Overall, some aspects did not affect the grade, whereas others were almost always considered when grading students. There were no differences between SP and faculty respondents regarding the quality of a good doctor. Concerning the qualities of a bad doctor, the SPs stressed more professional manners, time devoted to each patient, and waiting times than faculty respondents did. For personal experiences with doctors, there were no differences between SP and faculty respondents, and both had good and bad

experiences. Both faculty and SP examiners had higher expectations for students in higher semesters. Higher-semester students are informed that they will be graded more strictly than lower-semester students on multiple occasions. The possible influence of this on student motivation and OSCE performance needs to be explored in future studies. Several studies have shown that text anxiety and stress may impact student performance. In South Korea, certain noncognitive factors influence students' anxiety while taking the OSCE. Kim (20) suggested developing interventions to reduce student anxiety. These can also be considered in our institution. Concerning the influence of examination grade on student motivation, there were individual differences among respondents in both groups, but no overall differences between the two groups. The consequences of failure differed between faculty and SP examiners. SPs do not formally mentor students; only faculty members do, and they may develop a closer relationship with a mentee, which may account for the differences. The appearance of the students may have a greater influence on the SP respondents. The female respondents may have been stricter about the examinee's appearance. The discrepancies between SP and faculty examiners in grading were due to the different percentages of marks allotted to faculty and SP examiners and the different parameters they were evaluating.

Among the factors that do not affect grading are the fees students pay. SPs and faculty were unanimous in their decision not to consider the tuition fees paid by students when grading or giving feedback on their communication skills. The consequences of grading a student poorly, which may demotivate them or even cause them to fail the examination, were also not factors considered when grading. It is essential to produce good-quality future doctors so that they can grade students on the basis of their expectations. Many faculty members felt that feedback is more important than grades in motivating or demotivating students and providing ways to improve. The exam rules and regulations are another factor that does not affect grading. The faculty perspective was that rules and regulations needed to be in place, as they felt that they provided a fair and standard examination for all students and helped examiners stay focused on the task at hand.

There were a few factors that clearly affected grading. The scale of examinations has a detrimental effect on grading in terms of exhaustion and fatigue for assessors, both SPs and faculty. Faculty appeared more affected than SPs, as SPs have the option to be substituted by reserves. Faculty, however, expressed difficulty concentrating, especially when the exam lasted for more than a few hours, especially after lunch. A common theme brought up by some SPs and faculty alike, especially in large-scale examinations, was the comparison of student performance. Most tend to use the first student or first group of students as a grading benchmark and even revisit and change the marks for previous students if a subsequent student performs much better or worse. One faculty member mentioned that finding the first group of students is challenging to grade, as there is no reference point. In contrast, another SP mentioned that the first few students are "lucky", as they do not have a comparison. Another factor affecting communication skill grading is the student's semester or year. Higher-semester students were graded more strictly.

Other factors revealed differing perspectives from faculty and SPs. The roles of SPs or faculty at a particular station may affect their grading expectations. For example, one SP respondent noted that if the SP's role is that of an elderly person or someone from a low educational background, students are expected to adjust their communication accordingly. One faculty member also mentioned that being an expert in a particular specialty may lead to higher expectations of that skill, which may influence grading. The final factor that appeared to bring about mixed views was student characteristics. While all SP and faculty respondents agreed that ethnicity, age, and race do not affect marking, other factors, such as grooming,

attire, accents, and educational backgrounds, may do so. Some SPs reported marking students down if attire and grooming are extremely poor. In contrast, others mentioned not considering it, as they would mark it based on how the student communicates.

Faculty scores are generally lower than those provided by SPs. Faculty members assess a wider range of skills than SPs do, and their experience as clinicians may influence the grading, as evident from their perspectives expressed during the FGDs. We may need to reassess the role of SPs and explore ways to strengthen their position, thereby achieving greater congruence between SPs and faculty. Based on the results of this study, calibration exercises to standardise expectations between faculty and SPs can be considered. These should focus on factors to be considered when evaluating communication and physical examination skills and developing common ground on SP and faculty expectations of a good doctor. Workshops and training sessions to achieve this goal may be considered. SPs play a vital role in student assessment in examinations such as the United States Medical Licensing Examination (USMLE). A recent article mentioned that standardised patient-based performance assessment is now included in undergraduate and postgraduate medical education and is commonly used to evaluate both the technical and nontechnical skills necessary for the safe and effective practice of medicine (21). While we did not specifically examine “failure to fail”, both SP and faculty participants mentioned it. The importance of assessing and creating good doctors was repeatedly mentioned. If students do not meet the required standards, they should fail, and feedback should be provided to improve.

Strengths and Limitations

This study explored factors contributing to grading discrepancies between SPs and faculty. The limitations were that the data for both cohorts were obtained during the pandemic, with one cohort’s examination being carried out face-to-face and the other cohort’s examination being carried out virtually. This study focused only on communication skills and did not explore content-based skills, techniques, or professionalism. The OSCE examiners’ ratings of communication skills were measured at different scales at all stations except one. This could have affected the interrater agreement between the raters’ faculty and the SPs. Although respondents were allowed the freedom to explore topics related to the topic under discussion, we acknowledge the possibility that using some closed-ended guided questions in the FGD guide may have reduced the depth of responses.

CONCLUSION

This study determined the interrater agreement between faculty and SP ratings of OSCE communication skills and the factors they consider when assessing examinees. Faculty’s and SPs’ different perceptions of communication ability contributed to some disparities in grading. There is a need to consider such variations during the standardisation process to ensure that faculty and SPs recognise the expected communication standards.

The factors identified offer valuable insights into the aspects of communication that evaluators prioritise, which can inform how stations are designed and standardised before each examination. By addressing these factors in preexam training sessions, faculty and SP assessment have better-aligned expectations. However, the observed discrepancies do not necessarily indicate a need for correction but rather serve to prepare students for the range of communication styles and perspectives they will encounter in their professional practice.

Addressing these aspects during standardisation would provide a more uniform and transparent grading procedure while maintaining the authenticity of diverse patient viewpoints, making the test fair and representative of real-world clinical settings.

ACKNOWLEDGEMENTS

We extend our gratitude to all participants, including faculty and simulated patients, who volunteered to participate in this research. A special acknowledgement goes to Professor Ian G Wilson, Adjunct Professor at the IMU Centre for Education, for his valuable assistance in reviewing and copyediting the manuscript as a native English speaker.

ETHICS APPROVAL

Ethical approval was received from the IMU University JC Committee, Project ID: IMU579-2023. Written consent was obtained from all participants.

REFERENCES

1. Choudhari S, Umamaheswari TN. Objective structured clinical examination (OSCE): a review. *Res J Pharm Technol*. 2016;9(8):1257–9. <https://doi.org/10.5958/0974-360X.2016.00238.92>
2. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. 1979;13(1):39–54. <https://doi.org/10.1111/j.1365-2923.1979.tb00918.x>
3. Wallace J, Rao R, Haslam R. Simulated patients and objective structured clinical examinations: review of their use in medical education. *Adv Psychiatr Treat*. 2002;8(5):342–8. <https://doi.org/10.1192/apt.8.5.342>
4. MacLean S, Geddes F, Kelly M, Della P. Simulated patient training: using interrater reliability to evaluate simulated patient consistency in nursing education. *Nurse Educ Today*. 2018;62:85–90. <https://doi.org/10.1016/j.nedt.2017.12.024>
5. Cooper C, Mira M. Who should assess medical students' communication skills: their academic teachers or their patients? *Med Educ*. 1998;32(4):419–21. <https://doi.org/10.1046/j.1365-2923.1998.00223.x>
6. Ali K, Hanks S, Cockerill J. What is the value of simulated patient assessment in structured clinical examinations of undergraduate students, and how should it be used? *Eur J Dent Educ*. 2022;26(2):416–21. <https://doi.org/10.1111/eje.12703>
7. Chen JY, Chin WY, Tsang JP. How clinician examiners compare with simulated patients in assessing medical student empathy in a clinical exam setting. *Med Teach*. 2020;42(1):86–91. <https://doi.org/10.1080/0142159X.2019.1665635>
8. Cuevas-Nunez MC, Pulido MT, Harpe S, Stein AB, Lempicki K. Assessment of communication and physical exam skills: a comparison of students, faculty and standardised patients. *J Dent Educ*. 2022;86(7):853–62. <https://doi.org/10.1002/jdd.12892>

9. Khan AS, Shabi-ul-Hassan S. Comparison of assessment scores of candidates for communication skills in an OSCE, by examiners, candidates and simulated patients. *Creat Educ.* 2012;3(6):931. <https://doi.org/10.4236/ce.2012.326141>
10. Yousuf F, Yousuf N. Agreement between simulated patients and faculty: assessment of communication skills during objective structured clinical examination. *Pak J Med Sci.* 2019;35(6):1570. <https://doi.org/10.12669/pjms.35.6.1000>
11. Rothman AI, Cusimano M. A comparison of physician examiners', standardised patient', and communication experts' ratings of international medical graduates' English proficiency. *Acad Med.* 2000;75(12):1206–11. <https://doi.org/10.1097/00001888-200012000-00018>
12. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012;22(3):276–82. <https://doi.org/10.11613/bm.2012.031>
13. Cheng L, DeLuca C, Braund H, Yan W, Rasooli A. Teachers' grading decisions and practices across cultures: exploring the value, consistency, and construction of grades across Canadian and Chinese secondary schools. *Stud Educ Eval.* 2020;67:100928. <https://doi.org/10.1016/j.stueduc.2020.100928>
14. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Couns Psychother Res.* 2021;21(1):37–47. <https://doi.org/10.1002/capr.12360>
15. Liew SC, Dutta S, Sidhu JK, De-Alwis R, Chen N, Sow CF, Barua A. Assessors for communication skills: SPs or healthcare professionals? *Med Teach.* 2014;36(7):626–31. <https://doi.org/10.3109/0142159X.2014.899689>
16. Pawluk SA, Zolezzi M, Rainkie D. Comparing student self-assessments of global communication with trained faculty and standardised patient assessments. *Curr Pharm Teach Learn.* 2018;10(6):779–84. <https://doi.org/10.1016/j.cptl.2018.03.012>
17. Wright B, McKendree J, Morgan L, Allgar VL, Brown A. Examiner and simulated patient ratings of empathy in medical student final year clinical examination: are they useful? *BMC Med Educ.* 2014;14:1–8. <https://doi.org/10.1186/1472-6920-14-199>
18. Taşdelen Teker G, Odabaşı O. Reliability of scores obtained from standardised patient and instructor assessments. *Eur J Dent Educ.* 2019;23(2):88–94. <https://doi.org/10.1111/eje.12406>
19. Hodges B, Regehr G, Martin D. Difficulties in recognising one's own incompetence: novice physicians who are unskilled and unaware of it. *Acad Med.* 2011;76(10):87–9. <https://doi.org/10.1097/00001888-201110000-00010>
20. Kim KJ. Factors associated with medical student test anxiety in objective structured clinical examinations: a preliminary study. *Int J Med Educ.* 2016;7:424. <https://doi.org/10.5116/ijme.5835.99d7>
21. Gerzina HA, Stovsky E. Standardised patient assessment of learners in medical simulation. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2023.