A Cross-Sectional Study Analysing the Reliability of the Simulated Patient as an Assessor of Medical Students' Professionalism during Objective Structured Clinical Examination

Soumendra Sahoo[1], Rashmirekha Sahoo[2], Minoo Pothen[3], Saurabh Srivastava[4], Mila Nu Nu Htay[5]

[1]Centre for Transformative Learning, Manipal University College Malaysia, Melaka, MALAYSIA

[2]Centre for Foundation and General Studies, Manipal University College Malaysia, Melaka, MALAYSIA

[3]Department of Psychiatry, Manipal University College Malaysia, Melaka, MALAYSIA

[4]Department of Ophthalmology, Manipal University College Malaysia, Melaka, MALAYSIA

[5]Department of Community Medicine, Manipal University College Malaysia, Melaka, MALAYSIA

──────── **ABSTRACT** ────────

During the examination of professionals, the simulated patient (SP) can be the right person to assess the professionalism of examinees. This research analysed the efficiency of SP assessments of students' professionalism. Students' professionalism was assessed during an objective structured clinical examination (OSCE) by two independent assessors (the examiner and the SP). Professionalism was assessed using the previously validated Medical Students' Professionalism Assessment Scale. The scores provided by two assessors were subject to descriptive analysis, with mean, standard deviation, median, interquartile range, and minimum and maximum values derived. Since the rating was based on the ordinal scale, the agreement between the two assessors was analysed using quadratic weighted kappa statistics. Both SPs and examiners provided similar mean scores in their assessments. The highest difference was observed in relation to one statement on "student causing pain to the SP", in which the SP assessment mean score was 0.10, while the examiner assessment mean score was 0.48. In terms of agreement, the item "the student gave clear instructions before examining me" achieved moderate concordance (agreement). Although most of the items achieved fair concordance between the assessors, we recommend periodic training of SPs to improve their abilities as assessors for rating the professionalism shown by medical students.

Keywords: OSCE, Professionalism, Assessment, Simulated patient, Medical students

──────── **CORRESPONDING AUTHOR** ────────

Soumendra Sahoo, Center for Transformative Learning, Manipal University College Malaysia, Bukit Baru, 75150 Melaka, Malaysia

Email: soumendra.sahoo@manipal.edu.my

INTRODUCTION

There has been a marked change in the delivery of medical education from the conventional mode of training to competency-based training (1). During simulated patient (SP)–student encounters, the students learn more than just communication skills; their interactions with SPs contribute to their professional and personal identity development.

Performance-based assessment is consistent with outcome-based education (2), whereby learners can demonstrate their performance of tasks, approach to tasks and professionalism. Specifically, standardised patient-based performance assessments have advanced to include undergraduate and graduate medical education and are commonly used to evaluate both the technical and non-technical skills necessary for the safe and effective practice of medicine (3–5). However, one study showed that the professionalism scores for second-year podiatric medical students during an SP encounter varied significantly based on whether the scores were given by faculty or SPs (6).

The Accreditation Council for Graduate Medical Education (ACGME) Outcome Project identified professionalism as one of six general competencies required for residents of all specialities. The ACGME defines professionalism "as manifested through a commitment to carrying out professional responsibilities, adherence to ethical principles, and sensitivity to a diverse patient population" (7). The Medical Student's Professionalism Assessment Scale (MSPAS) is simple to navigate and can be completed quickly if SPs are provided the proper training on this assessment. The MSPAS was found to have good and reliable psychometric value (8). In another study, the Philadelphia College of Osteopathic Medicine (PCOM) measured the quality of osteopathic medical students' (Doctor of Osteopathis Medicine [DO]students) interpersonal and communication skills through SP encounters in which SPs rated the DO students relative to eight criteria (items), which constitute the Professionalism Assessment Rating Scale (PARS). These criteria are linked in the literature to patient outcomes, patient adherence, patient satisfaction and malpractice.

This study investigated the psychometric properties of the PARS based on data from the SP encounters of 205 osteopathic medical students of a PCOM class during their first three years of medical school (9). To the best of our knowledge, the comparison of tutor and SP ratings during an objective structured clinical examination (OSCE) to determine the efficiency of SPs as assessors of professionalism has not been reported. Hence, this study used the MSPAS to compare an SP's and an examiner's assessment of medical students' professionalism during the OSCE part of a final-year examination of undergraduate medical students.

MATERIALS AND METHODS

Samples

The sample comprised final-year medical students who undertook their professional examination at a private medical university in Malaysia. The sample was selected using a universal sampling method, recruiting from among those who sat for an OSCE. The sample of 60 students was assessed for professionalism by an examiner and an SP.

Assessors

One faculty examiner and one SP assessed the professionalism of each student throughout their OSCE. The students' professionalism was assessed at one of the OSCE stations, where a physical examination of the eye was the task. The professionalism assessment was done simultaneously and independently by those two assessors. Both the SP and the examiner attended two training sessions for professionalism assessment before the OSCE.

Assessment Tool

The professionalism of the students was assessed using a previously validated scale, the MSPAS (10). The scale included six items, and each item was assessed using a 10-point Likert scale (0 = strongly disagree, 10 = strongly agree).

Statistical Analysis

A descriptive analysis of the scores provided by the two assessors was carried out for each item, with mean, standard deviation, median, interquartile range, minimum and maximum values derived.

Reliability Analysis

The internal consistency of the physical examination professionalism assessment scale was assessed using Cronbach's alpha correlation coefficients. The alpha coefficient was 0.60 for the SP's assessment and 0.92 for the examiner's assessment.

Interrater Agreement

Since ratings were provided based on an ordinal scale, the agreement between the two assessors was analysed using quadratic weighted kappa statistics (11, 12). The concordance of the quadratic weighted kappa statistics (κ_w) was interpreted as follows: negative = poor, 0.01–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial and 0.81–1.00 = almost perfect. A value of 0.05 indicated statistical significance.

RESULTS

A total of 60 students were assessed by both the SP and the examiner. A descriptive analysis of the ratings is presented in Table 1. The median score was 10 for all items except for Item 5 (The student caused excessive/unnecessary pain while examining me), which was a negative item on the scale. Both the SP and the examiner provided similar mean scores in their assessments. The highest difference was observed for Item 5, for which the SP gave a mean score of 0.10, while the examiner gave a mean score of 0.48 (Table 1).

Table 1: Summary of scores provided by the SP and the examiner for students (n = 60)

Items	SP's assessment	Examiner's assessment
1. The student addressed me respectfully/politely		
Median (IQR)	10 (10, 10)	10 (10, 10)
Mean (SD)	9.93 (0.25)	9.8 (0.48)
Minimum, maximum	9, 10	8, 10
2. The student's questions were clear to me		
Median (IQR)	10 (10, 10)	10 (10, 10)
Mean (SD)	9.88 (0.32)	9.57 (0.89)
Minimum, maximum	9, 10	6, 10
3. The student gave clear instructions before examining me		
Median (IQR)	10 (9, 10)	10 (9, 10)
Mean (SD)	9.60 (0.67)	9.52 (0.89)
Minimum, maximum	7, 10	6, 10
4. The student was gentle in performing a physical examination on me		
Median (IQR)	10 (10, 10)	10 (9, 10)
Mean (SD)	9.77 (1.18)	9.43 (1.02)
Minimum, maximum	1, 10	6, 10
5. The student caused excessive/unnecessary pain while examining me		
Median (IQR)	0 (0, 0)	0 (0, 1)
Mean (SD)	0.10 (0.35)	0.48 (0.83)
Minimum, maximum	0, 2	0, 4
6. I would like to be examined by this student again		
Median (IQR)	10 (9, 10)	10 (9, 10)
Mean (SD)	9.62 (0.64)	9.48 (1.00)
Minimum, maximum	8, 10	6, 10

The ratings provided by the SP and the examiner were compared, and the findings are presented in Table 2. Item 3 (The student gave clear instructions before examining me), was linked to moderate agreement between the assessors (κ_w 0.58). For the other items, there was a fair agreement between the assessors (Table 2).

Table 2: Comparison of professionalism assessment scores given by the SP and the examiner for students (n = 60)

Item	Quadratic weighted Kappa (κ_w)	95% CI		Concordance
		Lower	Upper	
1. The student addressed me respectfully/ politely	0.35	0.03	0.67	Fair
2. The student's questions were clear to me	0.27	0.08	0.46	Fair
3. The student gave clear instructions before examining me	0.58	0.35	0.81	Moderate

(Continued on next page)

Table 2: (*Continued*)

Item	Quadratic weighted Kappa (κ_w)	95% CI		Concordance
		Lower	Upper	
4. The student was gentle in performing a physical examination on me	0.33	0.09	0.58	Fair
5. The student caused excessive/ unnecessary pain while examining me	0.28	−0.04	0.61	Fair
6. I would like to be examined by this student again	0.34	0.10	0.58	Fair

DISCUSSION

Definitions of professionalism often focus on attitudes and traits such as honesty, altruism, self-reflection, reliability and respect for others. The operationalisation of such abstract definitions is challenging, as they are subject to various interpretations when translated into measurable behaviours. Despite such challenges, specific behavioural benchmarks can be developed and utilised in evaluations using available methods, such as patient/nurse surveys, faculty observations, OSCEs, ethical reasoning tests and completion of administrative tasks (13).

In our study, both the examiner and the SP provided similar mean professionalism assessment scores for six descriptive items, except Item 5 (The student caused excessive/unnecessary pain while examining me). This may be because the SP was more accommodating and wanted to help students as the SPs knew it was these students' final exam. This finding made us reevaluate and revamp SP training. In another study, SP ratings were less reliable and consistent than physician or lay ratings, although the SPs ranked students more consistently than the other rater types (4). Although the scores provided by the SP and the examiner in our study achieved fair to moderate agreement, the level of agreement can be improved to reach substantial or almost-perfect concordance between the assessors. In the future, augmentation of SP training would be beneficial to achieve the desired level of agreement in professionalism assessments carried out by examiners and SPs concurrently.

Item 3 (The student gave clear instructions before examining me), achieved moderate concordance based on the weighted kappa score, while other items achieved fair concordance. This can be attributed to the fact that Item 3 is subjective in nature.

One study (14) found that SPs can give clinically meaningful assessment scores on empathy. However, in our study, there was a difference in the examiner's versus the SP's score for Item 5 (The student caused excessive/unnecessary pain while examining me). This might be because of the too-artificial context of the OSCE. Furthermore, the SP might not have been skilled enough to rate students.

This study was limited by the small sample size and the use of only one SP. Additionally, the OSCE station featuring communication tasks was not included. Furthermore, we did not analyse other factors that might have affected the results, such as the medical students' or assessors' experience, simulation training or perceptions. Further research with a much larger sample and the use of various clinical tasks can reveal more insights into the assessment of SPs.

CONCLUSION

Most professional assessment scores offered by the SP matched those offered by the examiner. Nevertheless, we recommend the periodic involvement of SPs in clinical teaching sessions and training them on how to assess professionalism to further improve their abilities as assessors of the professionalism shown by medical students.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the facilitation by the Centre for Transformative Learning, Manipal University College Malaysia exam cell and Vice Chancellor's office. This article was presented at the 7th International Conference on Research in Teaching and Education, Oxford, UK, on 25th August 2023.

ETHICAL APPROVAL

This study was approved by institutional research and ethics committee(Ref: MMMC/FOM/Research Ethics Committee -12/2018).

REFERENCES

1. Lurie SJ. History and practice of competency-based assessment. Med Educ. 2012;46(1):49–55. https://doi.org/10.1111/j.1365-2923.2011.04142.x

2. Harden RM, Crosby JR, Davis MH, riedman, M. AMEE guide no. 14: outcome-based education: part 5–From competency to meta-competency: a model for the specification of learning outcomes. Med Teach. 1999;21(6):546–52. https://doi.org/10.1080/01421599978951

3. Barrows HS. An overview of the uses of stan dardized patients for teaching and evaluating clinical skills. Acad Med. 1993;68(6):443–51. https://doi.org/10.1097/00001888-199306000-00002

4. Zanetti M, Keller L, Mazor K, Carlin M, Alper E, Hatem D, et al. Using standardized patients to assess professionalism: a generalizability study. Teach Learn Med. 2010;22(4):274–9. https://doi.org/10.1080/10401334.2010.512542

5. Winkel AF, Gillespie C, Uquillas K, Zabar S, Szyld D. Assessment of developmental progress using an objective structured clinical examination-simulation hybrid examination for obstetrics and gynecology residents. J Surg Educ, 2016;73(2):230–7. https://doi.org/10.1016/j.jsurg.2015.10.004

6. Park YS, Hyderi A, Heine N, May W, Nevins A, Lee M, et al. Validity evidence and scoring guidelines for standardized patient encounters and patient notes from a multisite study of clinical performance examinations in seven medical schools. Acad Med. 2017;92(11S):S12–20. https://doi.org/10.1097/ACM.0000000000001918

7. Williams BW, Byrne PD, Welindt D, Williams MV. Miller's pyramid core competency assessment: a study in relationship construct validity. J Contin Educ Health Prof. 2016;36(4):295–9. https://doi.org/10.1097/CEH.0000000000000117

8. Mahoney JM, Vardaxis V, Anwar N, Hagenbucher J. Differences in faculty and standardized patient scores on professionalism for second-year podiatric medical students during a standardized simulated patient encounter. J Am Podiatr Med Assoc. 2018;108(2):145–50. https://doi.org/10.7547/15-175

9. Accreditation Council for Graduate Medical Education (ACGME). Outcome project. Chicago, IL: ACGME/American Board of Medical Specialties (ABMS); 1999.

10. Htay MNN, Shrivastava S, Sahoo S. Development and validation of Medical Students' Professionalism Assessment Scale (MSPAS) in OSCE: simulated patients' rating scale. MedEdPublish. 2019;8:94. https://doi.org/10.15694/mep.2019.000094.1

11. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. Int J Nurs Stud. 2011;48(6):661–71. https://doi.org/10.1016/j.ijnurstu.2011.01.016

12. Abraham B, Nair MS. Automated grading of prostate cancer using convolutional neural network and ordinal class classifier. Inform Med Unlocked. 2019;17:100256. https://doi.org/10.1016/j.imu.2019.100256

13. Nichols BG, Nichols LM, Poetker DM, Stadler ME. Operationalizing professionalism: a meaningful and practical integration for resident education. Laryngoscope. 2014;124(1):110–5. https://doi.org/10.1002/lary.24184

14. 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;14:199. https://doi.org/10.1186/1472-6920-14-199