Res Dev Med Educ 2026;15:33367
doi:10.34172/rdme.33367
https://rdme.tbzmed.ac.ir

Original Article

Reliability of Peer and Self-assessment Across learning domains in Competency-Based Medical Education for Formative Assessment: An Observational Study

Sangeetha Kandasamy[1*], Thilakavathi Kothandan[1], Shivkumar Gopalakrishnan[2], Yogambal Muthureddy[3], Vijay Anto James[4], Ganesh Kumar Puthisigamani[1]

[1]Government Villupuram Medical College & Hospital, Villupuram, Tamil Nadu, India
[2]Government Medical College & Hospital, Orathur, Nagapattinam, Tamil Nadu, India
[3]Government Stanley Medical College, Chennai, Tamil Nadu, India
[4]The Gandhigram institute of rural health (MoHFW), Dindigul, Tamil Nadu, India

Article info
Article History:
Received: July 22, 2025
Revised: September 21, 2025
Accepted: October 6, 2025
epublished: July 14, 2026

Keywords:
Self-assessment, Peer assessment, Learning domains, Formative assessment, Reliability, CBME

Abstract

Introduction: In Competency-Based Medical Education, formative assessments are integrated throughout the curriculum to offer continuous feedback, promoting self-regulatory learning and helping learners to identify the areas for improvement. Peer and self-assessment play a crucial role in CBME by fostering collaboration, self-reflection, critical thinking, and professionalism. Assess the reliability of peer and self-assessment across different learning domains in a CBME framework for formative assessment.

Methods: This observational study, conducted from January 2024 to December 2024 at a Government Tertiary Care Hospital, involved a universal sampling method with 100 second-year medical students. Assessments addressed knowledge through assignments, psychomotor skills through vitals and systemic exams, and affective skills through communication and interpersonal abilities. Evaluations by peers, self, and faculty were conducted using an internally validated tool, with anonymity maintained throughout.

Results: Participants were aged 18–21 with a male-to-female ratio of 1:1.2. ANOVA tests showed no significant difference in knowledge scores between self, peer, and faculty assessments (ICC = 0.708; $P = 0.106$). Significant differences were found in psychomotor ($P = 0.001$; ICC = 0.483) and attitude ($P = 0.001$; ICC = 0.234) domains, with students and peers overestimating or underestimating performance. Bland–Altman plots confirmed good agreement for knowledge, moderate for psychomotor, and poor for attitude between self, peer, and faculty ratings.

Conclusion: The study found moderate reliability in the knowledge domain but poor to no reliability in the attitude and psychomotor domains, with self and peer assessments tending to overestimate performance. These findings underscore the need for more training and structured assessment strategies to enhance the accuracy and consistency of self- and peer-assessment in medical education

Introduction

Over the last few decades, medical education has evolved significantly, with the 1970s and 1980s marking a shift toward reflective practice, experiential learning, and self-regulation. These theories laid the foundation for advanced assessment methods such as self-assessment (SA) and peer-assessment (PA), which emphasise clear benchmarks to guide professional growth. Today, PA and SA are integral to multisource feedback and assessment for learning, promoting professionalism and continuous competency development in medical education.[1] While informal training in earlier times occurred through mentoring and experiential learning, reduced learner interaction and evolving educational needs have created a greater demand for structured, formal assessments like SA and PA.

PA and SA are increasingly recognised as vital components of medical education, particularly in India. These methods encourage self-reflection, critical thinking, and accountability.[2] With the introduction of Competency-Based Medical Education (CBME), formative assessments are embedded across the

***Corresponding author:** Sangeetha Kandasamy, Email: sangeetha6482@gmail.com

curriculum, offering ongoing feedback and aligning outcomes with healthcare needs.[3,4] Unlike traditional time-based training, CBME emphasises the achievement of specific competencies in a student-centred model.

These assessments introduce a collaborative framework where students learn from one another, share diverse perspectives, and develop constructive feedback skills.[5,6] This interaction reduces hierarchy, fosters camaraderie, and enhances communication skills essential for patient care.[7] PA also reinforces accountability and ethical standards by allowing students to observe and evaluate peers' professional behaviours in clinical settings.[8,9] However, some hesitate to participate due to concerns about fairness or relationships. Supportive environments, structured guidelines, and feedback training can mitigate these barriers[5,7-10]When embedded into curricula, SA prepares students for autonomous practice and lifelong competency development.[11,12] A meta-analysis of 175 studies (19,383 participants) showed that SA (g = 0.585) and PA (g = 0.606) significantly enhance academic performance, underscoring their role in competency-based medical education. [13]

Using PA and SA across cognitive, psychomotor, and affective domains ensures a holistic evaluation of medical students. In the cognitive domain, students identify knowledge gaps, strengthen analytical skills, and link theory with practice. In the psychomotor domain, PA and SA provide opportunities to evaluate clinical skills, build confidence, and practice hands-on learning. Structured rubrics and faculty guidance are essential to ensure reliability, as students often feel uneasy about evaluating technical skills. The affective domain particularly benefits from peer evaluation, as students can closely observe professionalism, communication, and interpersonal behaviours. This complements faculty evaluation by reinforcing ethics, empathy, and accountability.

For example, studies report that medical students engaging in PA and SA improved their problem analysis, identified learning needs, and strengthened collaborative knowledge sharing.[9,10] Professionalism is most often evaluated within the affective domain, as peers are well-positioned to observe attitudes and behaviours across varied settings.[2,14] Many students find peer evaluation of psychomotor skills valuable, but may feel uncomfortable providing critical feedback. Traditional peer roles and anxiety about fairness contribute to this hesitation. However, studies suggest that early exposure and faculty supervision improve both comfort and reliability.[14–17] Students prefer anonymous, immediate feedback focused on professional and unprofessional behaviours.

Despite concerns about fairness and objectivity, peer and self-assessment remain powerful educational strategies. They foster critical skills, enhance professionalism, and bridge theoretical knowledge with practical application.[14,16] When integrated across learning domains, these assessments promote deeper learning, reflection, and competency development. PA and SA address cognitive, psychomotor, and affective outcomes simultaneously.[18] Together, they support holistic evaluation, encourage critical thinking, and ensure that graduates of CBME are competent, ethical, and collaborative medical professionals. [19]

Research question

What is the reliability of peer and self-assessment across learning domains of knowledge, skill, and affective in CBME compared with faculty assessment?

Purpose of the study

While many studies have examined either self-assessment or peer assessment, most have focused on a single learning domain. This study aims to explore the utility of both peer and self-assessment across all three learning domains within medical education, even though the scope is limited to medical students.

Objectives

- To develop an internally validated assessment tool for peer and self-assessment in Knowledge, Skill, and Affective communication domains.
- To implement and analyze the effectiveness of peer assessment and self-assessment with faculty assessment in various domains for the second-year Medical students.

Methods

Study design and setting

This observational study was conducted from January 2024 to December 2024 at a Government Tertiary Care Hospital by the Medical Education Unit, involving second-year medical students. Participants were assessed across three learning domains: knowledge, psychomotor, and affective, through self, peers & faculty evaluations.

Study Participants and Sampling

A universal sampling method was used, involving 100 second-year Medical students enrolled in the Competency-Based Medical Education program course, who were eligible and hence all were included.

Intervention and evaluation plan

The proposed CBME curriculum for the vertical integration module on anaemia links the departments of Paediatrics and Pathology. It addresses all three learning domains using context-rich methods grounded in CBME principles, Harden's integration ladder, and Miller's pyramid (Table 1). The design aligns Harden's integration levels with Miller's assessment levels to promote a holistic, learner-centred approach reflecting real clinical challenges. For the cognitive domain, lectures introduced the pathophysiology, types, causes, and management of anaemia. The assessment used a validated 10-mark rubric that covered depth, relevance, clarity, and accuracy. This reflects a multidisciplinary approach (Harden level 5) and the "knows–knows how" levels of Miller's pyramid. Clinical competencies were developed

Table 1. Anaemia integration framework across learning domains with assessment tools, methods, and validation in a CBME context

Domain	Competencies	Harden's Integration	Miller's pyramid level	Assessment Method	Assessment tool & validation	Mode of evaluation
Knowledge	Disease understanding: anaemia	Multidisciplinary level-5 shared teaching by pathology and pediatrics	Knows how	Structured written assignment	Internally validated structured 5-point rubric (content, depth, accuracy, relevance, clarity) Content validity index≥0.85 with discipline experts	Self, Peer, & faculty
Psychomotor	Clinical examination & Haemoglobin estimation skills	Interdisciplinary level -9- 10 Integrated clinical teaching using a patient-centred linked case	Shows how	Bedside clinical skill demonstration during postings	Objective structured checklist plus global rating scale. Rubric piloted and revised; Interrater reliability, Cohen's d>0.75	Faculty, Peer observer, Self-reflective checklist
Affective	Therapeutic management: Communication &Counselling	Interdisciplinary level -9- 10 Authentic real case scenario & Simulation setup	Shows how	Clinical encounters	Checklist based on the Kalamazoo communication model.Face validity by a senior expert panel.	Self, Peer, & Faculty

through demonstrations and real patient interactions in postings, involving communication and counselling with caregivers. This represents interdisciplinary to transdisciplinary learning (Harden levels 9–10). The assessment utilised a simulation with an OSCE checklist based on the Kalamazoo consensus framework, aligning with the "shows how" level of Miller's model. To ensure quality and fairness, assessments were jointly prepared by subject experts, the alignment and integration committee, and the Medical Education Unit members.

A structured three-step validation process was followed for the assessment rubrics, beginning with content validation. This was carried out by a panel of experts using the Delphi method to ensure the relevance and clarity of each item. The 30-item questionnaire was designed and rated by six speciality experts, with a minimum threshold value of 0.86 set for inclusion in the final tool.

After content validation, a pilot study was conducted with 20 students from another medical college to check feasibility, language clarity, and inter-rater agreement. Finally, the framework was evaluated across domains, which confirmed strong internal consistency for each domain-specific knowledge ($\alpha \geq 0.86$), psychomotor OSCE ($\alpha \geq 0.82$), and affective ($\alpha \geq 0.8$). Reliability was tested using the Test-Retest method using the same assessment tool after a week's interval.

Each student was assessed by themselves, two peers, and one faculty member, whose results were averaged within learning domains. Peer assessment was randomly allocated using a lottery method independently. Because the integrated anaemia CBME module involved both pathology and paediatrics disciplines, based on expertise, varied across the domains; however, every student received two peer raters and one faculty rater per domain. In reliability, raters were considered a random sample of domain-specific assessors. Interrater reliability was measured using a two-way random effects, average measure, agreement intraclass correlation coefficient.

Data collection tool and technique

During the intervention, data collection was done under strict ethical and confidentiality guidelines. Study participants' evaluations were recorded anonymously using alphanumeric codes to maintain identity and eliminate assessor bias. Assessors (faculty, peers, and self) were trained to use the tools objectively. Evaluation of learning outcomes using structured, validated checklists to assess the knowledge gain, skill development, and communication proficiency. All assessments were conducted in a blinded, anonymous manner to reduce bias. All data were stored securely and were only accessible to core research personnel for analysis.

Statistical method

Comparison of repeated measures scores among more than two groups is analysed using repeated measures ANOVA or the Friedman test. A five percent level of significance is considered statistically significant. The Intraclass Correlation Coefficient (ICC) was used to evaluate the reliability of scores between students, peers, and faculty across the three assessed domains: knowledge, skill, and attitude. Statistical analysis is performed using Microsoft Excel and the R package. A Bland-Altman plot was used to evaluate the agreement between self-assessed and peer-assessed scores. The reliability of the numerical scores of repeated measures is assessed using the intraclass correlation coefficient (ICC).

Results

A total of 100 students participated in the study. The participants were aged between 18 and 21, and the male-to-female ratio was 1:1.2. The comparison of mean +/-SD scores in the domains of knowledge, skill, and attitude, as assessed by students, peers, and faculty, is depicted in Table 2.

Knowledge domain

There was no significant difference in knowledge scores between self-assessment, peer assessment, and faculty assessment. This suggests reasonable agreement across the three groups in evaluating knowledge (Figure 1, 2 & 3).

Psychomotor domain

A significant difference was observed in psychomotor skill scores between self-assessment, peer assessment, and faculty assessment. Both self and peer scores were higher compared to faculty scores, indicating a tendency for students and peers to overestimate psychomotor skills relative to faculty evaluation (Figures 1, 2 & 3).

Affective domain

There were also considerably high scores in self and

Table 2. Comparing the knowledge, skill, and attitude scores of self, peers, and faculty

Domain	Group	Mean ± SD	*P*-value	Significance
Knowledge	Self-score	8.76 ± 0.80	0.106	Not significant
	Peer-score	8.99 ± 0.92		
	Faculty score	8.84 ± 1.25		
Psychomotor skill	Self-score	6.21 ± 0.84	0.001**	Significant
	Peer-score	6.17 ± 0.59		
	Faculty score	5.07 ± 1.15		
Attitude	Self-score	4.00 ± 1.00*	0.001**	Significant
	Peer-score	4.00 ± 1.00*		
	Faculty score	4.00 ± 0.00		

* Median (IQR). & ***P*-value ≤ 0.05 is considered statistically significant.

peer assessment for attitudes and communication skills as compared with faculty assessment. This potentially reflects subjective biases or differing interpretations of the criteria (Table 3 & Figure 2 & 3).

Comparison between types of assessments.

Self and peer assessment

A Bland-Altman plot was used to evaluate the agreement between self-assessed and peer-assessed scores, showing good agreement in the scores for knowledge, psychomotor skills, and attitude. This showed that self and peers are on the same page with regard to the assessment of all 3 domains.

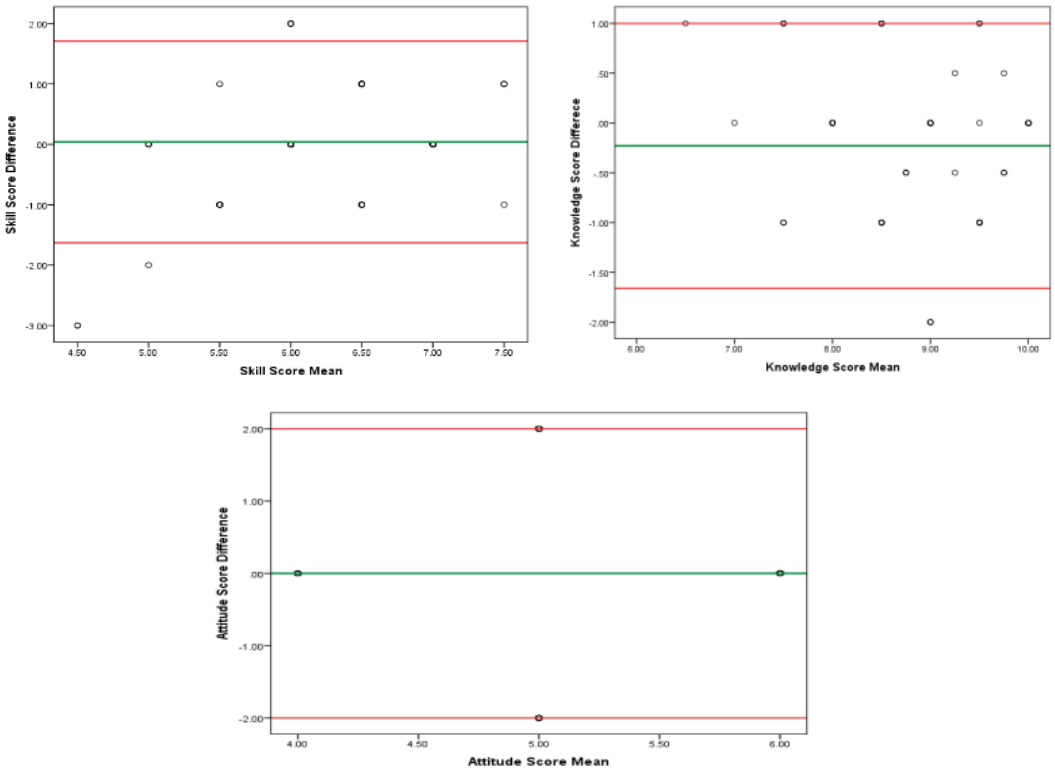

Figure 1. Bland–Altman plot comparing self-assessment and peer assessment for knowledge, skills, and attitude

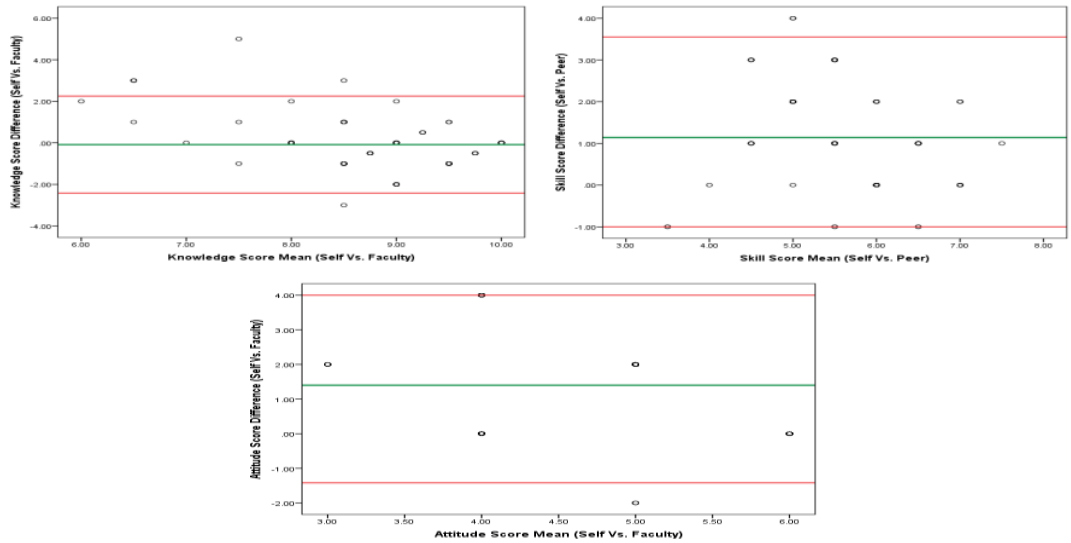

Figure 2. Bland–Altman plot comparing self-assessment and faculty assessment for knowledge, skills, and attitude

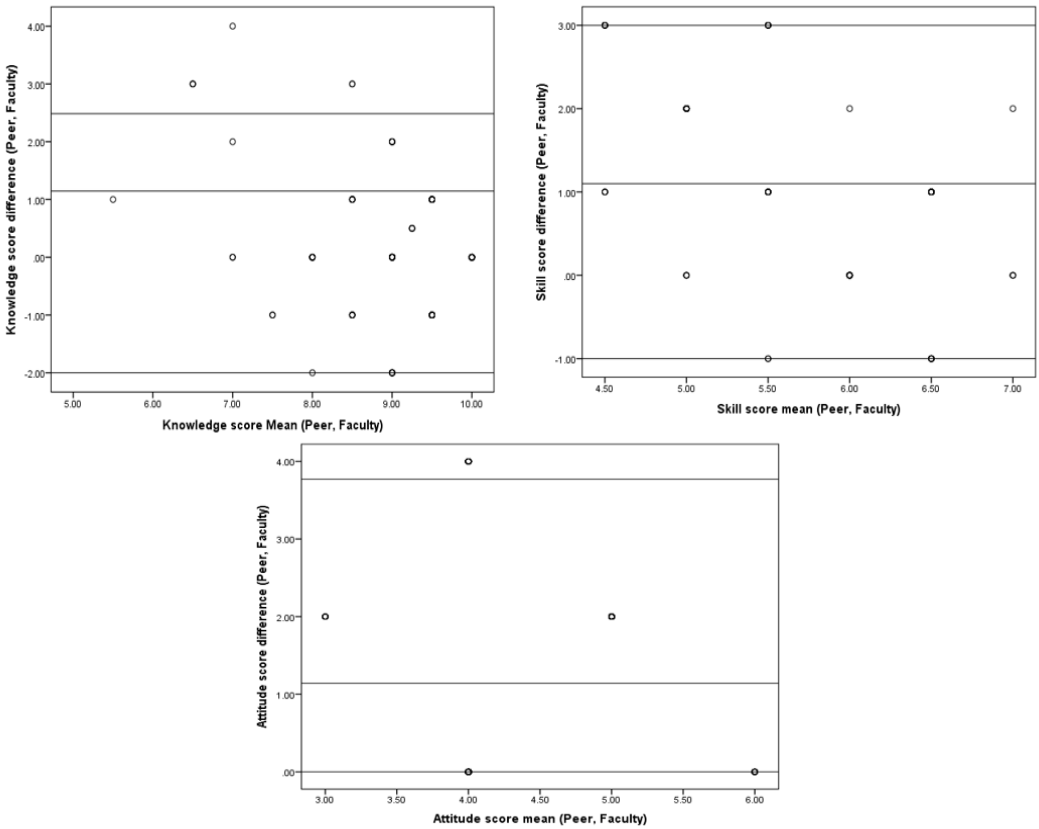

Figure 3. Bland–Altman plot comparing peer assessment and faculty assessment for knowledge, skills, and attitude

Table 3. Intraclass Correlation Coefficient (ICC) between the scores of Self, Peers, and Faculty

Domain	ICC (Intraclass Correlation)	95% CI (Lower - Upper)	*P*-value	Reliability Interpretation
Knowledge	0.708	0.593 - 0.795	0.001	Moderate reliability
Skill	0.483	0.278 - 0.636	0.001	Poor reliability
Attitude	0.234	-0.069 - 0.461	0.059	Non-reliable

Self and faculty assessment

The plot demonstrated good agreement between self-assessed and faculty-assessed scores for knowledge and attitude, with moderate agreement observed for psychomotor skill scores (Figures 1, 2 & 3). This observation could evoke 2 opinions:

1. Whether it is faculty or student self-assessment, knowledge and attitude scores are the same. This means both assessors are following standard guidelines and checklist-based scoring, which results in minimal variation in scores for these 2 domains.
2. When assessing for psychomotor skills, student self-assessment seems to be more liberal. The faculties are more objective in their assessment of psychomotor skills, and hence the scores were moderately lower. Students seem to overestimate their psychomotor skills.

Peer and faculty assessment

Peer assessment seemed to match faculty assessment in the knowledge domain. This might be due to the same reasons, such as standardised questionnaire, checklist, and prior sensitisation of peers in assessing knowledge. However, peers put high scores for students in the psychomotor and affective skill domains. Faculty assessment scores were much lower in these areas, which is probably due to less bias, more stringent observation, and objective scoring methods.

The Intraclass Correlation Coefficient (ICC) was used to evaluate the reliability of scores between students, peers, and faculty across the three assessed domains: knowledge, skill, and attitude. The ICC for the knowledge domain was 0.708 (95% CI: 0.593–0.795, $P=0.001$), indicating moderate reliability. This suggests an acceptable level of agreement between students, peers, and faculty in assessing knowledge-based assignments. The ICC for the skill domain was 0.483 (95% CI: 0.278–0.636, $P=0.001$), reflecting poor reliability. The students and peers overrate their performance. This indicates limited agreement between the raters in assessing psychomotor skills, such as clinical examinations. The ICC for the attitude domain was 0.234 (95% CI: -0.069–0.461, $P=0.059$), signifying no reliability (Table 3). Here again, the peers and students have a false high perception of their skills, whereas the faculty feel otherwise. This suggests that there was minimal or no agreement between students, peers, and faculty in evaluating attitudes and communication skills.

Discussion

Peer assessment is a common practice in medical education, while self-assessment became more prominent after the adoption of CBME and serves as an effective

approach to peer teaching and learning among medical students. It is utilised both as an assessment tool and a learning strategy, fostering social accountability within the learning environment. These assessments help prepare students for self-monitoring and self-regulation, which are essential for lifelong learning, while also promoting active participation in the classroom. Research has shown that peer assessment is effective in improving knowledge acquisition in medical and health sciences education.[18] A recent meta-analysis confirmed that both self- and peer-assessment significantly enhance academic performance across disciplines and contexts.[14] In clinical and procedural settings, peer assessment can be used to evaluate practical skills such as surgical techniques, physical examinations, or communication skills in simulations. Studies suggest that peer assessment in psychomotor domains fosters skill acquisition through observation, feedback, and reflection.[19,20] Peer assessment also has a significant impact on the affective domain, particularly in developing interpersonal communication, empathy, and professional attitudes. When peers assess each other's communication skills or teamwork, it encourages self-awareness and emotional intelligence. Moreover, peer feedback helps in understanding the perspectives of others, which is a critical component of affective learning. [21-23] The most recognised purpose of peer assessment is its use as an assessment tool. To ensure the judgments made by peers are valid and reliable, specific quality criteria can be established based on the objectives of peer assessment as an evaluative method.[23,24] This aligns with findings in nursing education, where self- and peer-evaluations tended to score higher than teacher evaluations, reinforcing their role as complementary rather than standalone methods for evaluating clinical performance. [25]

Research indicates that the reliability of peer assessments is often at or above acceptable levels. Peer assessment can vary in reliability, influenced by several factors. While many studies suggest that peer assessments can achieve acceptable levels of reliability and validity, the consistency of ratings can fluctuate based on implementation. Reliability varies significantly depending on contextual factors such as the educational setting, task complexity, and the characteristics of the peer raters involved.[27,28] This study demonstrated that while peer and self-assessment showed moderate reliability in the knowledge domain, the reliability was poor for psychomotor skills and nil for attitudes and communication. This underscores the need for improving assessment tools and providing additional training to enhance the consistency and reliability of evaluations in skill-based and affective domains. Students and peers tended to overestimate psychomotor and affective scores, highlighting variability and subjective biases. The educational background of participants can affect their assessment capabilities. Providing clear and detailed rubrics can enhance the reliability of peer assessments by ensuring that all participants understand the expectations and standards for evaluation.[27,29] Students often require explicit training in assessment techniques to improve their ability to provide high-quality feedback. Without adequate training, assessments may be influenced by biases or inconsistencies, leading to lower reliability.[26,30,31] Some studies have found that peer assessments can yield relatively high reliability when compared to expert evaluations. However, these research works were done in controlled settings and reflect short-term outcomes with inherent limitations.[28,32] This suggests that while peer assessment is valuable, it should be complemented with instructor evaluations for more reliable outcomes. Addressing these factors can help mitigate issues related to reliability in peer assessment.

In our study, the CBME-integrated module was effective in assessing the knowledge domain with structured and validated tools demonstrating consistent results across assessors. However, greater variability and subjectivity in psychomotor and affective assessments highlight the need for improved assessor training and refinement of evaluation tools to ensure reliability in clinical and communication competencies.

Limitation

This research work was a single-centre study conducted with a universal sampling method for second-year medical students and may restrict the generalizability of the findings to broader academic contexts. Time constraints allowed assessing only one integrated topic, with a focus on psychometric analysis primarily in a formative setting. In addition, limited rater training and potential subjectivity in self and peer assessments may have influenced the reliability, particularly in the psychomotor and affective domains.

Conclusion

Peer assessment and self-assessment have often achieved acceptable levels of reliability and validity in previous studies, but they can show variability influenced by factors such as context, clarity of criteria, and the training provided to students. It can achieve reliable performance provided these factors are adequately addressed. Student training measures such as structured orientation sessions, use of standardised rubrics, practice assessments with feedback, and calibration exercises can drive peer assessment to near-perfection levels. The results emphasise the need for such strategies to improve the accuracy and consistency of peer and self-assessment in medical education.

Acknowledgments

The authors acknowledge the institutional administration for their support and the students for their voluntary participation.

Authors' Contribution

Conceptualisation: Thilakavathi Kothandan, Sangeetha Kandasamy.
Data Curation: Sangeetha Kandasamy, Thilakavathi Kothandan.
Investigation: Sangeetha Kandasamy, Thilakavathi Kothandan.
Methodology: Sangeetha Kandasamy, Vijay Anto James.
Project Administration: Sangeetha Kandasamy.
Resources: Sangeetha Kandasamy, Shivkumar Gopalakrishnan.
Writing–Original Draft: Sangeetha Kandasamy, Shivkumar Gopalakrishnan.

Writing–Review & Editing: Shivkumar Gopalakrishnan, Yogambal, Ganesh Kumar Puthisigamani

Competing Interests
The authors declare no conflict of interest.

Ethical Approval
This study was approved by the Institutional Ethics Committee (IEC) with the code of the institution (GVMC/IEC/2023(2)/2).

Funding
None.

References
1. Abraham R, Singaram VS. Self and peer feedback engagement and receptivity among medical students with varied academic performance in the clinical skills laboratory. BMC Med Educ 2024;24(1):1065. doi:10.1186/s12909-024-06084-9
2. Surapaneni KM. Importance of peer and patient feedback in facilitating and assessing the learning progression of medical students. Med Sci Educ 2023;33(5):1283. doi:10.1007/s40670-023-01844-6
3. Shah N, Desai C, Jorwekar G, Badyal D, Singh T. Competency-based medical education: an overview and application in pharmacology. Indian J Pharmacol 2016;48(Suppl 1):S5-9. doi:10.4103/0253-7613.193312
4. Bhattacharya S. Competency-based medical education: an overview. Ann Med Sci Res 2023;2(3):132-8. doi:10.4103/amsr.amsr_27_23
5. Arnold L, Shue CK, Kritt B, Ginsburg S, Stern DT. Medical students' views on peer assessment of professionalism. J Gen Intern Med 2005;20(9):819-24. doi:10.1111/j.1525-1497.2005.0162.x
6. Biesma R, Kennedy MC, Pawlikowska T, Brugha R, Conroy R, Doyle F. Peer assessment to improve medical student's contributions to team-based projects: randomised controlled trial and qualitative follow-up. BMC Med Educ 2019;19(1):371. doi:10.1186/s12909-019-1783-8
7. Sharma A, Dada R. Peer assessment: an effective tool to modify professional behaviour of undergraduate medical students. Indian J Clin Anat Physiol 2023;10(3):136-37. doi:10.18231/j.ijcap.2023.030
8. Khorwal G, Meshram R, Vaibhav V, Sharma R, Singh B, Chandran S, et al. Peer assessment of student presentations: key takeaways and lessons learned. Cureus 2024;16(5):e59809. doi:10.7759/cureus.59809
9. Double KS, McGrane JA, Hopfenbeck TN. The impact of peer assessment on academic performance: a meta-analysis of control group studies. Educ Psychol Rev 2020;32(2):481-509. doi:10.1007/s10648-019-09510-3
10. Tayem YI, James H, Al-Khaja KA, Razzak RL, Potu BK, Sequeira RP. Medical students' perceptions of peer assessment in a problem-based learning curriculum. Sultan Qaboos Univ Med J 2015;15(3):e376-81. doi:10.18295/squmj.2015.15.03.012
11. Maas MJ, Sluijsmans DM, van der Wees PJ, Heerkens YF, Nijhuis-van der Sanden MW, van der Vleuten CP. Why peer assessment helps to improve clinical performance in undergraduate physical therapy education: a mixed methods design. BMC Med Educ 2014;14:117. doi:10.1186/1472-6920-14-117
12. Zheng B, He Q, Lei J. Informing factors and outcomes of self-assessment practices in medical education: a systematic review. Ann Med 2024;56(1):2421441. doi:10.1080/07853890.2024.2421441
13. Kaur M, Singh H, Badyal D. Impact of continuous self-assessment on the learning of medical students. Int J Res Med Sci 2016;4(7):2665-9. doi:10.18203/2320-6012.ijrms20161929
14. Yan Z, Lao H, Panadero E, Fernández-Castilla B, Yang L, Yang M. Effects of self-assessment and peer-assessment interventions on academic performance: a meta-analysis. Educ Res Rev 2022;37:100484. doi:http://dx.doi.org/10.1016/j.edurev.2022.100484
15. Curran VR, Fairbridge NA, Deacon D. Peer assessment of professionalism in undergraduate medical education. BMC Med Educ 2020;20(1):504. doi:10.1186/s12909-020-02412-x
16. Alzaabi S, Nasaif M, Khamis AH, Otaki F, Zary N, Mascarenhas S. Medical students' perception and perceived value of peer learning in undergraduate clinical skill development and assessment: mixed methods study. JMIR Med Educ 2021;7(3):e25875. doi:10.2196/25875
17. Engelmann JM. Peer assessment of clinical skills and professional behaviors among undergraduate athletic training students. Athl Train Educ J 2016;11(2):95-102. doi:10.4085/110295
18. Capan Melser M, Lettner S, Bäwert A, Puttinger C, Holzinger A. Pursue today and assess tomorrow - how students' subjective perceptions influence their preference for self- and peer assessments. BMC Med Educ 2020;20(1):479. doi:10.1186/s12909-020-02383-z
19. Topping K. Peer assessment between students in colleges and universities. Rev Educ Res 1998;68(3):249-76. doi:http://dx.doi.org/10.3102/00346543068003249
20. Lerchenfeldt S, Taylor TAH. Best practices in peer assessment: training tomorrow's physicians to obtain and provide quality feedback. Adv Med Educ Pract 2020;11:571-8. doi:10.2147/amep.S250761
21. Boud D, Falchikov N. Rethinking Assessment in Higher Education: Learning for the Longer Term. 1st ed. London: Routledge; 2007. doi:10.4324/9780203964309
22. van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. Med Teach 2012;34(3):205-14. doi:10.3109/0142159x.2012.652239
23. Wisniewski B, Zierer K, Hattie J. The power of feedback revisited: a meta-analysis of educational feedback research. Front Psychol 2019;10:3087. doi:10.3389/fpsyg.2019.03087
24. Norcini JJ. Peer assessment of competence. Med Educ 2003;37(6):539-43. doi:10.1046/j.1365-2923.2003.01536.x
25. Ghaljeh M, Rezaee N, Arbabisarjou A. Comparison of self-, peer, and teachers' evaluation about the clinical skills of nursing students at the department of psychiatry. J Educ Health Promot 2021;10:397. doi:10.4103/jehp.jehp_1034_20
26. Speyer R, Pilz W, Van Der Kruis J, Brunings JW. Reliability and validity of student peer assessment in medical education: a systematic review. Med Teach 2011;33(11):e572-85. doi:10.3109/0142159x.2011.610835
27. Lin Z, Yan H, Zhao L. Exploring an effective automated grading model with reliability detection for large-scale online peer assessment. J Comput Assist Learn 2024;40(4):1535-51. doi:10.1111/jcal.12970
28. Power JR, Tanner D. Peer assessment, self-assessment, and resultant feedback: an examination of feasibility and reliability. Eur J Eng Educ 2023;48(4):615-28. doi:10.1080/03043797.2023.2185769
29. Xiong Y, Schunn CD, Wu Y. What predicts variation in reliability and validity of online peer assessment? A large-scale cross-context study. J Comput Assist Learn 2023;39(6):2004-24. doi:10.1111/jcal.12861
30. Sluijsmans DM, Brand-Gruwel S, van Merriënboer JJ. Peer assessment training in teacher education: effects on performance and perceptions. Assess Eval High Educ 2002;27(5):443-54. doi:10.1080/0260293022000009311
31. Gielen M, De Wever B. Peer assessment in a wiki: product improvement, students' learning and perception regarding peer feedback. Procedia Soc Behav Sci 2012;69:585-94. doi:10.1016/j.sbspro.2012.11.450
32. Lerchenfeldt S, Mi M, Eng M. The utilization of peer feedback during collaborative learning in undergraduate medical education: a systematic review. BMC Med Educ 2019;19(1):321. doi:10.1186/s12909-019-1755-z