GBMN GEORGIAN BIOMEDICAL & MEDICAL NEXUS
Former Georgian Biomedical News
VOLUME 4 ISSUE 3. JUL-SEP 2026

# Structured Peer-Evaluation Training in OSCE-Based Clinical Skills Education: A Non-Randomized Controlled Pilot Study

Gvantsa Vardosanidze[1,ID], Nino Shiukashvili[1,ID], Nino Tevzadze[1,ID], Mariam Rochikashvili[1,ID], Krishna Rajendra Abnawe[1,ID], Sharath Penneru[1], Sonal Sharanappa Purad[1,ID], Abdur Raheem Mohammed[1,ID], Eka Ekaladze[1,ID]

## ABSTRACT

**BACKGROUND.** Structured peer-evaluation is increasingly used as a formative strategy in medical education, but its value depends on structured implementation and appropriate student preparation. In clinical skills training, peer assessment may help students engage more actively with performance criteria, feedback, and professional behavior.

**OBJECTIVES.** This pilot study explored whether structured peer-evaluation training was associated with changes in fourth-semester medical students' OSCE-based clinical skills performance, observed professional behavior, and alignment of peer–instructor assessments.

**METHODS.** A prospective controlled pre–post pilot educational study was conducted at Ken Walker International University during one academic semester. Twenty-one fourth-semester medical students participated: 11 in the experimental group and 10 in the control group. The experimental group received structured peer-evaluation and feedback training after the midterm OSCE, while the control group received faculty assessment only. Students were assessed at midterm and final OSCEs using standardized clinical skills checklists and a professional behavior checklist. In the experimental group, peer and instructor ratings were also compared. Scores were converted into percentages and analyzed using exploratory pre–post and between-group comparisons.

**RESULTS.** In the experimental group, faculty-assessed professional behavior improved from a median of 86.7% to 93.0%, while physical examination performance improved from 82.0% to 94.0%. The control group showed smaller, non-significant changes. However, between-group comparisons of change scores were not statistically significant. Peer ratings of professional behavior showed a stronger positive association with instructor ratings after training, whereas peer ratings of physical examination performance remained weakly associated with instructor ratings.

**CONCLUSIONS.** Structured peer-evaluation training was feasible and associated with favorable within-group changes in professional behavior and physical examination performance. Findings suggest that students may assess observable professional behaviors more accurately than technical clinical skills. Because of the small sample size, non-random allocation, baseline imbalance, and variation in OSCE stations across time points, the results should be interpreted as preliminary and hypothesis-generating.

**KEYWORDS.** Clinical skills education; Medical education; Objective Structured Clinical Examination (OSCE); Peer evaluation; Professional behavior; Undergraduate medical students.

DOI. 10.52340/GBMN.2026.01.01.181

## BACKGROUND

Clinical skills education is a core component of undergraduate medical training, requiring students to develop procedural competence, communication skills, professional behavior, and the ability to reflect on performance. Objective Structured Clinical Examinations (OSCEs) are widely used to assess clinical performance because they allow structured observation of defined competencies under standardized conditions. In addition to their assessment role, OSCEs can serve a formative function by providing feedback that helps students identify performance gaps and prepare for subsequent clinical encounters or examinations.

Formative OSCE-based activities may be particularly valuable when they create opportunities for students to observe performance, compare it with explicit criteria, and receive or provide feedback. Group-based and formative OSCE models have shown that structured self-, peer-, and tutor feedback can be integrated into clinical skills assessment to support learning prior to subsequent examinations. [1] Similarly, peer and self-assessment have been explored in formative OSCE settings, particularly for communication and clinical performance skills, suggesting that learners can benefit from being actively involved in the assessment process rather

than remaining passive recipients of faculty judgment alone. [2]

Peer assessment has therefore been proposed as a complementary strategy in undergraduate clinical skills education. By assessing classmates using structured checklists, students may become more aware of expected performance standards, common errors, and the observable features of effective clinical practice. Studies comparing peer and faculty assessments in clinical skills and simulation-based settings suggest that peer assessment can contribute to learning and may show meaningful alignment with instructor assessment in some performance domains. [3,4] However, the accuracy of peer assessment varies across clinical tasks and contexts.

This inconsistency is especially important in technical skills assessment. While students may be able to observe visible communication and professional behaviors, they may have more difficulty judging procedural accuracy, sequencing, and completeness of physical examination steps. Prior research on peer and faculty assessment of advanced clinical skills has shown that agreement may be limited and that peer assessment requires careful preparation before use for educational purposes. [5] Therefore, structured peer evaluation should be implemented as a structured, supervised, formative activity rather than as an informal substitute for faculty assessment.

Professional behavior is another important component of clinical skills performance. Behaviors such as respectful communication, permission-seeking, active listening, appropriate patient positioning, empathy, and avoidance of unnecessary discomfort are observable during simulated clinical encounters. Because these behaviors are visible and behavior-based, they may be suitable for structured peer observation and feedback. [6] At the same time, assessment of professional behavior requires clear criteria to reduce subjectivity and support consistency between evaluators. [7]

Although peer assessment has been explored in OSCE and clinical skills education, further evidence is needed on how structured peer-evaluation training functions in undergraduate clinical skills courses. In particular, there is a need to examine whether peer-evaluation training is associated with changes in OSCE-based performance, observed professional behavior, and alignment between peer and instructor ratings. This is especially relevant in educational contexts where structured peer-evaluation may provide an additional supervised formative learning opportunity without replacing faculty assessment.

To address this gap, the present pilot study implemented a structured peer-evaluation and feedback-training intervention among fourth-semester medical students during an OSCE-based clinical skills course. The intervention introduced constructive feedback principles and structured feedback models, including the Situation-Behavior-Impact and feedback sandwich models. Students who received peer-evaluation training and peer assessment in addition to standard faculty assessment were compared with students who received faculty assessment only.

This study aimed to explore whether structured peer-evaluation training was associated with changes in fourth-semester medical students' OSCE-based clinical skills performance, observed professional behavior, and alignment of peer-instructor assessments. Given the pilot nature of the study, the findings were intended to be exploratory and hypothesis-generating rather than confirmatory.

## METHODS

This study was designed as a prospective, controlled, pre–post pilot educational study conducted at Ken Walker International University over one academic semester. It was considered exploratory and hypothesis-generating rather than confirmatory. The study explored whether structured peer-evaluation training was associated with changes in clinical skills performance and observed professional behavior among fourth-semester medical students.

Fourth-semester medical students enrolled in the same clinical skills course were recruited through convenience sampling. Participation was voluntary, and informed consent was obtained. A total of 21 students participated: 11 students in the experimental group and 10 students in the control group.

### Peer-evaluation and feedback training

Students were assessed at two time points: a midterm OSCE in April 2025 and a final OSCE in June 2025. Both

the midterm and final OSCEs included a cardiovascular examination station. The stations were designed to be of equivalent difficulty, and the same checklist structure and maximum score were used at both time points. The peer-feedback training was conducted for the experimental group after the midterm OSCE, in late April. The training was delivered by faculty facilitators/clinical mentors and focused on providing clear, specific, constructive, respectful, and behavior-focused feedback.

The session introduced two structured feedback models: the Situation–Behavior–Impact model and the feedback sandwich model. Training activities included a short presentation, role-play demonstration of effective and ineffective feedback, guided practice in pairs or small groups, faculty coaching, and feedback tip sheets with example phrases. Students in the experimental group subsequently practiced giving peer feedback before the final OSCE.

## OSCE procedure

At both the midterm and final OSCEs, students were assessed on physical examination and history-taking using the standard KWIU OSCE checklists. Professional behavior was assessed concurrently during the clinical encounter. Standardized patients were used at both examinations. The same faculty examiners assessed students at both time points and were blinded to students' group assignment throughout the study.

In the control group, students were assessed only by faculty examiners. In the experimental group, students were assessed by faculty examiners and also by trained peer evaluators. Peer evaluators independently observed classmates and completed structured peer-evaluation forms. Peer-evaluation scores and feedback were used only for research and educational purposes and did not affect official grades.

## Assessment tools

Clinical skills performance was assessed using standard KWIU OSCE checklists for physical examination and history-taking. Faculty examiners used these checklists for students in both groups, while peer evaluators used them only in the experimental group. The same checklist structure and maximum possible score were maintained across the midterm and final OSCEs to ensure consistent assessment. Scores were converted into percentages for comparison.

Professional behavior was assessed by faculty examiners using a 10-item professional behavior checklist. The checklist assessed observable behaviors including punctuality, appropriate positioning in relation to the patient, explanation of the examination purpose in simple language, asking permission before touching or entering personal space, open body posture, preparedness with required equipment, active listening and allowing time for patient interaction, appropriate eye contact, avoidance of unnecessary discomfort, and maintenance of a calm, confident, and empathetic demeanor. Each item was rated on a three-point scale: 1=does not meet expectations, 2=satisfactory, and 3=excellent. Total scores were converted into percentage scores.

Peer feedback quality was assessed by the teacher/mentor using a separate Peer Feedback Quality Checklist. This checklist assessed ten feedback-quality dimensions: clarity, specificity, constructiveness, positivity/balance, relevance, objectivity, respectful and professional tone, use of a structured feedback framework such as Sandwich or SBI, active observation of the peer, and usefulness for improvement. Each criterion was rated as "Yes," "Partially," or "No," with space for comments.

## Data collected

The collected data included faculty-assessed physical examination, history-taking, and professional behavior scores for both groups at the midterm and final OSCEs. In the experimental group, peer-assessed clinical skills scores were also collected. All checklist scores were recorded using anonymized codes and converted into percentages where applicable.

## Data analysis

Checklist scores were converted to percentages before analysis. For the experimental group, raw clinical-skill scores were converted by dividing the obtained score by the maximum checklist score and multiplying by 100. Professional behavior scores were converted using a maximum score of 30 points. Control group scores were already recorded as

percentages. The main pre–post analysis included 21 students: 11 in the experimental group and 10 in the control group. One additional control student appearing in the final checklist was excluded from the main pre–post analysis because no corresponding pre-test data were available.

Descriptive statistics were summarized as medians and interquartile ranges (IQRs) for all outcomes. Owing to the small sample size and non-normal distribution of the data, non-parametric methods were used throughout the analyses. No adjustment for multiple comparisons was applied because of the exploratory nature of the study. Given the small sample size, the results were interpreted primarily as exploratory trends, supported by effect sizes, rather than as definitive evidence of intervention effectiveness.

Within-group changes from midterm to final OSCE were evaluated using the Wilcoxon signed-rank test for paired samples. Between-group differences in change scores (post-test minus pre-test) were assessed using the Mann–Whitney U test. Baseline and post-test comparisons between the experimental and control groups were also performed using the Mann–Whitney U test. Effect sizes for between-group comparisons were reported as rank-biserial correlation coefficients. The association between peer and faculty ratings was examined using Spearman's rank-order correlation coefficient (rho), and paired differences between peer and faculty ratings were evaluated using the Wilcoxon signed-rank test. All statistical tests were two-tailed, and statistical significance was defined as $p<0.05$.

### Ethical considerations

Participation was voluntary, and informed consent was obtained from all students. Student identities were protected through anonymized coding. Peer-evaluation scores and peer-feedback activities were used only for educational and research purposes and did not affect official grades. Faculty supervised the peer evaluation and feedback process to ensure fairness, confidentiality, and appropriate feedback practices. Ethical approval was obtained from the Ken Walker International University Institutional Review Board (IRB approval number: #1-2026/001).

## RESULTS

Faculty-assessed professional behavior improved in the experimental group from a median of 86.7% on the pre-test to 93.0% on the post-test. The median change was +10.0 percentage points, and the within-group change was statistically significant. In the control group, professional behavior changed from 68.3% to 66.0%, with a median change of +3.0 percentage points; this within-group change was not statistically significant.

Faculty-assessed physical examination performance also improved in the experimental group, increasing from a median of 82.0% at pre-test to 94.0% at post-test. The median change was +11.3 percentage points and was statistically significant. In the control group, physical examination scores increased from 63.5% to 64.5%, with a median change of +2.9 percentage points; this change was not statistically significant (**TAB.1**).

**TABLE 1**. Faculty-assessed pre–post changes in professional behavior and physical examination scores

| Outcome | Group | Pre-test median [IQR] | Post-test median [IQR] | Change median [IQR] | Within-group p-value |
|---|---|---|---|---|---|
| Professional behavior | Experimental | 86.7 [80.0-86.7] | 93.0 [88.0-97.0] | 10.0 [1.3-13.2] | 0.042 |
| Professional behavior | Control | 68.3 [58.4-72.5] | 66.0 [63.0-71.2] | 3.0 [-3.0-15.2] | 0.416 |
| Physical examination | Experimental | 82.0 [75.5-85.7] | 94.0 [91.0-97.0] | 11.3 [6.4-15.3] | 0.003 |
| Physical examination | Control | 63.5 [48.7-75.5] | 64.5 [61.2-74.5] | 2.9 [-7.3-15.4] | 0.375 |

Abbreviations: IQR, interquartile range.

Although the experimental group showed statistically significant within-group improvement, between-group comparison of change scores did not reach statistical significance. For professional behavior, the experimental group improved by a median of 10.0 percentage points compared with 3.0

percentage points in the control group. For physical examination, the experimental group improved by a median of 11.3 percentage points compared with 2.9 percentage points in the control group. However, the difference in change scores between groups was not statistically significant for either outcome (**TAB.2**).

**TABLE 2**. Between-group comparison of change scores

| Outcome | Experimental change median [IQR] | Control change median [IQR] | Mann–Whitney U p-value | Rank-biserial effect size |
|---|---|---|---|---|
| Professional behavior | 10.0 [1.3-13.2] | 3.0 [-3.0-15.2] | 0.526 | 0.17 |
| Physical examination | 11.3 [6.4-15.3] | 2.9 [-7.3-15.4] | 0.275 | 0.29 |

Abbreviations: IQR, interquartile range.

## Baseline imbalance

The two groups were not equivalent at baseline. The experimental group had higher pre-test scores than the control group in both professional behavior and physical examination. This baseline imbalance limits causal interpretation and supports treating the findings as exploratory rather than confirmatory (**TAB.3**).

**TABLE 3**. Baseline and post-test comparison between groups

| Outcome | Time point | Experimental median [IQR] | Control median [IQR] | Between-group p-value | Rank-biserial effect size |
|---|---|---|---|---|---|
| Professional behavior | Pre-test | 86.7 [80.0-86.7] | 68.3 [58.4-72.5] | <0.001 | 0.91 |
| Professional behavior | Post-test | 93.0 [88.0-97.0] | 66.0 [63.0-71.2] | <0.001 | 0.91 |
| Physical examination | Pre-test | 82.0 [75.5-85.7] | 63.5 [48.7-75.5] | 0.012 | 0.65 |
| Physical examination | Post-test | 94.0 [91.0-97.0] | 64.5 [61.2-74.5] | <0.001 | 0.95 |

Abbreviations: IQR, interquartile range.

Because baseline scores were already higher in the experimental group, post-test differences alone should not be interpreted as evidence of intervention effectiveness. The change-score analysis is more appropriate for this design.

## Peer versus instructor assessment in the experimental group

Peer and instructor ratings were compared within the experimental group. At pre-test, peer ratings of professional behavior were slightly higher than instructor ratings, but the difference was not statistically significant. At post-test, peer ratings of professional behavior became slightly lower than instructor ratings, with a median peer–instructor difference of -3.0 percentage points. The post-test correlation between peer and instructor professional behavior ratings was very strong.

For physical examination, peer and instructor scores were similar at pre-test. At post-test, peers rated physical examination performance lower than instructors, with a median peer–instructor difference of -8.0 percentage points. However, the correlation between peer and instructor physical examination scores remained weak and non-significant (**TAB.4**).

**TABLE 4**. Peer–instructor comparison in the experimental group

| Outcome | Time point | Instructor median [IQR] | Peer median [IQR] | Peer–instructor median [IQR] | Wilcoxon p-value | Spearman rho | Spearman p-value |
|---|---|---|---|---|---|---|---|
| Professional behavior | Pre-test | 86.7 [80.0-86.7] | 90.0 [81.7-93.3] | 3.3 [-5.0-10.0] | 0.391 | -0.60 | 0.051 |
| Professional behavior | Post-test | 93.0 [88.0-97.0] | 90.0 [84.5-95.0] | -3.0 [-3.5-0.0] | 0.047 | 0.96 | <0.001 |
| Physical examination | Pre-test | 82.0 [75.5-85.7] | 85.2 [73.8-88.3] | 4.1 [-4.9-4.9] | 0.898 | 0.30 | 0.373 |
| Physical examination | Post-test | 94.0 [91.0-97.0] | 88.0 [84.0-91.0] | -8.0 [-10.0- -4.0] | 0.008 | 0.26 | 0.432 |

Abbreviations: IQR, interquartile range.

## DISCUSSION

This prospective controlled pre–post pilot study explored whether structured peer-evaluation training was associated with changes in clinical skills performance, observed professional behavior, and peer-assessment alignment among fourth-semester medical students. The peer-evaluation intervention was associated with statistically significant within-group improvements in faculty-assessed professional behavior and physical examination performance from pre-test to post-test. In contrast, the control group showed smaller, statistically non-significant changes. However, between-group comparisons of change scores were not statistically significant. Therefore, the findings should be interpreted as preliminary and hypothesis-generating rather than as definitive evidence of intervention effectiveness.

The favorable trends observed within the experimental group are consistent with the broader educational rationale for peer assessment as a formative learning strategy. Structured peer evaluation requires students to actively observe performance, compare observed behaviors with explicit criteria, and provide feedback to colleagues. These processes may promote feedback literacy, self-regulated learning, and evaluative judgment, particularly when peer assessment is implemented using structured criteria and guided training. [8,9] In the present study, students in the experimental group were exposed not only to faculty assessment but also to peer-evaluation activities, which may have encouraged closer attention to checklist criteria and observable clinical behaviors.

The favorable trend in faculty-assessed professional behavior is particularly relevant. Professional behavior in this study was assessed through observable actions, including communication clarity, permission-seeking, appropriate positioning, active listening, respectful interaction, avoidance of unnecessary patient discomfort, and maintenance of a calm and empathetic demeanor. These behaviors are visible during clinical encounters and may therefore be especially suitable for peer observation and feedback. Previous work has suggested that peer assessment may be useful in professionalism-related domains because peers can provide complementary perspectives on behaviors observed during educational activities. [10] However, professionalism assessment remains methodologically challenging because definitions, tools, and assessment approaches vary across medical education contexts. [11] Accordingly, the present findings support the feasibility of structured observation of professional behavior but do not validate the checklist as a comprehensive instrument for assessing professionalism.

Faculty-assessed physical examination performance also improved in the experimental group from the midterm to the final OSCE. In this study design, the midterm OSCE served not only as a baseline assessment but also as a formative checkpoint, allowing students to receive feedback and identify areas for improvement before the final examination. Therefore, the observed improvement may reflect the educational value of intermediate feedback, checklist-guided peer observation, and


ISSN (Online): 2720-8796 ISSN (Print): 2720-7994

opportunities for students to reflect on expected examination sequences and performance standards. Acting as a peer evaluator may have shifted students from passive participation to active analytical observation, helping them recognize both correct performance and common omissions and apply these insights to subsequent clinical skills performance. However, this interpretation should be approached with caution because the midterm and final OSCEs assessed different clinical stations. As a result, improvement may also reflect normal course progression, increased OSCE familiarity, differences in station difficulty, or repeated exposure to assessment criteria in addition to the peer-evaluation intervention.

The peer-instructor comparison adds an important layer to the interpretation. After training, peer ratings of professional behavior showed a strong positive association with instructor ratings, suggesting that students may be able to judge observable professional behaviors more accurately after structured guidance and supervised peer-evaluation practice. This finding is consistent with previous work indicating that peer and near-peer feedback can be educationally valuable when students are prepared for the feedback role and when the process is structured and supportive. [12,13] In contrast, peer ratings of physical examination performance remained weakly correlated with instructor ratings. This contrast suggests that professional behavior, because it includes visible interpersonal and communication behaviors, may be more accessible to student raters than technical procedural performance, which requires procedural knowledge, correct sequencing, clinical accuracy, and recognition of missed examination steps.

These findings highlight the importance of structured feedback training before implementing peer evaluation. The intervention introduced feedback principles and structured models, including the Situation-Behavior-Impact and feedback sandwich models. This aligns with recommendations that peer assessment should not be implemented as an informal activity; students need explicit preparation in how to observe, judge, phrase, and deliver feedback constructively. [14] Feedback literacy is increasingly recognized as an important component of health professions education because students need to understand, interpret, and use feedback for improvement. [15] In the present study, structured training may have contributed to improved peer calibration in the professional behavior domain, although the study design does not allow this mechanism to be confirmed.

From a practical perspective, structured peer evaluation may offer an additional supervised formative learning opportunity in settings where faculty time and assessment resources are limited. Recent work on peer-feedback tools in clinical skills education has emphasized the potential scalability of structured peer feedback while also noting concerns about validity, variable engagement, and the need for supervision. [16] In the present study, peer scores did not affect official grades and were used only for educational and research purposes. This was appropriate given the pilot nature of the study and the limited evidence regarding the reliability of peer assessment, particularly for technical clinical skills.

## Limitations

Several limitations should be acknowledged. First, the sample size was small, which limited statistical power and increased the risk of unstable estimates. Second, students were allocated to existing tutorial groups rather than randomized individually, which may have introduced selection bias. Third, the experimental group had higher baseline scores than the control group, limiting causal interpretation of post-test differences. Therefore, the observed within-group improvements in the experimental group should be interpreted cautiously, and the findings should be considered preliminary. In addition, baseline demographic and educational characteristics were not collected, limiting the ability to assess baseline comparability between groups. Fourth, although the quality of peer feedback was assessed using a structured checklist, formal psychometric validation of the tool was not conducted. Finally, the study was conducted at a single institution, limiting generalizability.

## Strengths and future directions

Despite these limitations, the study has several strengths. It used a prospective pre-post design with a comparison group, incorporated structured peer-

feedback training, used checklist-based assessment, separated faculty assessment from peer assessment, and examined both student performance and peer-assessment alignment. The study also included practical safeguards such as anonymized coding, non-friend peer assignment, faculty supervision, and non-summative use of peer scores. These features support the feasibility of implementing structured peer evaluation as a supervised formative activity in undergraduate clinical skills education.

Future studies should include larger samples, randomized or matched group allocation where feasible, and comparable OSCE stations across pre- and post-assessments. Future research should also include formal peer-rater calibration, multiple faculty raters, inter-rater reliability analysis, and validation of peer-feedback quality tools. Qualitative analysis of written or oral peer feedback comments would also be valuable for examining whether students' feedback becomes more specific, constructive, balanced, and behavior-focused after training. Repeated peer-evaluation activities across several semesters may provide stronger evidence on whether peer assessment contributes to sustained development of clinical skills, professional behavior, and evaluative judgment.

## CONCLUSIONS

This pilot study suggests that structured peer-evaluation training is feasible and was associated with favorable within-group changes in faculty-assessed professional behavior and physical examination performance. Peer ratings of professional behavior showed a stronger positive association with instructor ratings after training, whereas peer ratings of physical examination performance remained weakly associated with instructor ratings. Because of the small sample size, non-random allocation, baseline imbalance, and variation in OSCE station content, the findings should be interpreted as preliminary and hypothesis-generating. Overall, the results support further investigation of structured peer evaluation as a supervised, formative educational strategy in undergraduate clinical skills training.

## AUTHOR AFFILIATION

1. Department of Basic Sciences, Ken Walker International University, Tbilisi, Georgia

## REFERENCES

1. Sulaiman ND, Shorbagi SI, Abdalla NY, Daghistani MT, Mahmoud IE, Al-Moslih AM. Group OSCE (GOSCE) as a formative clinical assessment tool for pre-clerkship medical students at the University of Sharjah. J Taibah Univ Med Sci. 2018;13(5):409-414. doi:10.1016/j.jtumed.2018.06.003.
2. Talwalkar JS, Murtha TD, Prozora S, Fortin AH VI, Morrison LJ, Ellman MS. Assessing advanced communication skills via objective structured clinical examination: A comparison of faculty versus self, peer, and standardized patient assessors. Teach Learn Med. 2020;32(3):294-307. doi:10.1080/10401334.2019.1704763.
3. 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.
4. Yu JH, Lee MJ, Kim SS, et al. Assessment of medical students' clinical performance using high-fidelity simulation: Comparison of peer and instructor assessment. BMC Med Educ. 2021;21(1):506. doi:10.1186/s12909-021-02952-w.
5. Moineau G, Power B, Pion AMJ, Wood TJ, Humphrey-Murto S. Comparison of student examiner to faculty examiner scoring and feedback in an OSCE. Med Educ. 2011;45(2):183-191. doi:10.1111/j.1365-2923.2010.03800.x.
6. Hoffman LA, Shew RL, Vu TR, Brokaw JJ, Frankel RM. The association between peer and self-assessments and professionalism lapses among medical students. Eval Health Prof. 2017;40(2):219-243. doi:10.1177/0163278717702191.
7. Setyonugroho W, Kennedy KM, Kropmans TJB. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. Patient Educ Couns. 2015;98(12):1482-1491. doi:10.1016/j.pec.2015.06.004.
8. Kim KJ, Kim G. The efficacy of peer assessment in objective structured clinical examinations for formative feedback: A preliminary study. Korean J Med Educ. 2020;32(1):59-65. doi:10.3946/kjme.2020.153.
9. Lerchenfeldt S, Taylor TAH. Best practices in peer assessment: Training tomorrow's physicians to obtain and provide quality feedback. Adv Med Educ Pract. 2020;11:571-578. doi:10.2147/AMEP.S250761.
10. 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.
11. Jaramillo Villegas C, Hernandez Rincon EH, Gallego Beltran JC, Guzman Sabogal YR. Assessment of professionalism in undergraduate and graduate medical education: A scoping review. BMC Med Educ. 2025;25(1):1295. doi:10.1186/s12909-025-07756-w.
12. Sader J, Cerutti B, Meynard L, et al. The pedagogical value of near-peer feedback in online OSCEs. BMC Med Educ. 2022;22:572. doi:10.1186/s12909-022-03629-8.
13. Schwill S, Fahrbach-Veeser J, Moeltner A, et al. Peers as OSCE assessors for junior medical students—a review of routine use: A mixed methods study. BMC Med Educ. 2020;20(1):17. doi:10.1186/s12909-019-1898-y.
14. Stenberg M, Mangrio E, Bengtsson M, Carlson E. Formative peer assessment in higher healthcare education programmes: A scoping review. BMJ Open. 2021;11(2):e045345. doi:10.1136/bmjopen-2020-045345.
15. Mohd Noor MN, Fatima S, Grace Cockburn J, Romli MH, Pallath V, Hong WH, Vadivelu J, Foong CC. Systematic review of feedback literacy instruments for health professions students. Heliyon. 2024;10(10):e31070. doi:10.1016/j.heliyon.2024.e31070.
16. Yurdal MO, Kıncal RY. Development of a digital peer-feedback tool for clinical skills training: A pilot study with second-year medical students. Sci Rep. 2025;15(1):40807. doi:10.1038/s41598-025-24650-z.