

Assessing Clinico-Social Case Discussions Among Undergraduate Medical Students: A Comparative Study of Structured and Conventional Methods

Jeny M. Mundodan ¹, S. S. Roshni ¹, Anupama Arumadi ¹, Minu Mohan ²

¹. Community Medicine, Government Medical College Manjeri, Manjeri, IND ². Community Medicine, Government Medical College, Ernakulam, Ernakulam, IND

Corresponding author: Jeny M. Mundodan, jeny.ali@gmail.com

Received 09/18/2026

Review began 09/21/2026

Review ended 10/22/2026

Published 10/24/2026

© Copyright 2026

Mundodan et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118516

Abstract

Background: Clinico-social case discussions are an important component of undergraduate Community Medicine practical assessment, evaluating students' ability to integrate clinical findings with social determinants, epidemiological reasoning, and preventive strategies. Conventional assessment methods allow holistic evaluation but may be affected by variability in examiner judgement. Structured assessment tools may improve clarity and consistency while preserving the comprehensive nature of case discussions.

Objectives: To compare the scoring patterns and inter-rater reliability of a structured assessment tool with the conventional method for clinico-social case discussions among Phase III MBBS students; and to describe faculty feedback regarding the feasibility and usefulness of the structured assessment approach.

Methods: A comparative observational study was conducted among 78 Phase III MBBS students posted in Community Medicine. A structured assessment tool with criterion-based domains and predefined scoring allocations was compared with the conventional assessment method. Thirty-two students were assessed simultaneously using conventional and structured methods to compare scoring patterns. Forty-six students were assessed by two examiners using the same method to determine inter-rater reliability. A supplementary descriptive component documented faculty feedback regarding the feasibility and usefulness of structured assessment using an open-ended questionnaire. Responses were summarized descriptively.

Results: Structured assessment scores were significantly lower than conventional scores (mean difference 1.31, $p < 0.001$), with moderate correlation between methods ($r = 0.605$). Both methods demonstrated good inter-rater reliability, with intraclass correlation coefficients (ICCs) of 0.782 and 0.826 for conventional and structured assessment, respectively. Faculty feedback highlighted the potential of structured assessment to improve clarity and standardization, while emphasizing the need to retain flexibility.

Conclusion: Structured assessment can provide a useful framework for clinico-social case discussions by improving transparency and guiding examiner judgement. Effective implementation requires balancing structured criteria with flexibility, along with examiner training.

Categories: Family/General Practice, Medical Education, Preventive Medicine

Keywords: community medicine, educational assessment, inter-rater reliability, structured assessment, undergraduate medical education

Introduction

Competency-Based Medical Education (CBME) emphasizes the demonstration of competencies required for effective clinical practice rather than the acquisition of factual knowledge alone. Accordingly, assessment should evaluate clinical reasoning, decision-making, communication skills, and application of knowledge in real-life contexts [1]. A good assessment system depends on factors such as sampling, implementation, and context of use, while balancing key attributes like validity, reliability, feasibility, acceptability, and educational impact [2,3].

Although traditional clinical examinations, such as long cases, evaluate integrated performance in realistic settings, they have been criticized for inconsistent scoring arising from variability in examiner expectations, questioning patterns, and case complexity, often due to the absence of definite assessment criteria [4]. Structured formats such as the Objective Structured Clinical Examination [5,6], Mini-Clinical Evaluation Exercise (Mini-CEX) [7], and the Objective Structured Long Examination Record (OSLER) [8] were developed to improve reliability through direct observation and criterion-referenced scoring. OSLER, in particular, retains the authenticity of the long case while incorporating predefined assessment domains and structured scoring [8]. However, findings on their effectiveness remain mixed, and examiner training is essential for acceptable reliability [9,10].

How to cite this article

Mundodan J M, Roshni S S, Arumadi A, et al. (October 24, 2026) Assessing Clinico-Social Case Discussions Among Undergraduate Medical Students: A Comparative Study of Structured and Conventional Methods. Cureus 18(10): e118516. DOI 10.7759/cureus.118516

Clinico-social case discussion in Community Medicine differs from conventional clinical assessment, requiring integration of clinical findings with epidemiological reasoning, social determinants of health, family and environmental factors, and community-based interventions. Despite their importance in undergraduate medical education, these assessments still rely largely on conventional methods, with considerable inter-examiner and inter-institutional variation, as existing structured assessment tools developed primarily for clinical disciplines may not adequately capture the public health competencies required in Community Medicine. While tools based on Mini-CEX principles have been pilot-tested for postgraduates [11], published evidence on structured assessment tailored for undergraduate clinico-social discussions remains limited.

Hence, there is a need for structured assessment approaches incorporating both clinical and public health dimensions for clinico-social case discussions in Community Medicine. This study aimed to compare the scoring patterns and inter-rater reliability of a structured assessment tool with the conventional method for clinico-social case discussions among Phase III undergraduate medical students. We also assessed faculty feedback regarding the feasibility and usefulness of the structured assessment approach.

Materials And Methods

We conducted a comparative cross-sectional observational study from April to July 2026 at Government Medical College Manjeri in Kerala, India. MBBS students in Phase III Part 1 underwent a four-week clinical rotation in Community Medicine in five consecutive batches. The study comprised two assessment components: a within-student comparison of structured and conventional assessment methods in the first two batches and an assessment of inter-rater reliability for each method in the subsequent three batches. A supplementary descriptive component was included to document faculty feedback regarding the feasibility and usefulness of the structured assessment method.

As this was a feasibility-based study conducted during scheduled undergraduate clinical postings, we did not perform a formal a priori sample size calculation. The number of participants was determined by the feasibility of simultaneous assessment by two examiners, with a pre-specified maximum of 16 students per batch. All 105 eligible students from five consecutive posting batches during the study period were invited to participate. Consecutive sampling was used, with students enrolled consecutively until this predefined batch-specific cap was reached, resulting in 78 participants. Two batches (Batches 3 and 4) contributed 15 participants each because the remaining eligible students were absent on the scheduled assessment day. Of the 105 eligible students, 14 were absent, six declined participation, and seven could not be included because the pre-specified recruitment cap had been reached in their respective batches. Recruitment was consecutive and was not based on academic performance, demographic characteristics, or examiner preference.

A structured case assessment tool for clinico-social case discussion was developed with contributions from two faculty members in Community Medicine. The tool used a 25-point scoring scale comprising multiple assessment domains, each with predefined scoring allocations and criterion-based scoring. The conventional method currently used involved unstructured case discussion with a global rating by the examiner based on overall student performance on the same 25-point scale, without predefined domain-wise scoring criteria. Six broad clinico-social case categories commonly used for undergraduate Community Medicine examinations at our institution were selected. Non-communicable diseases and communicable diseases were represented as broad categories for practical reasons, as the core assessment domains and scoring criteria were similar across conditions within these categories. Topic-specific structured assessment tools were then developed. The assessment tool was reviewed by three medical education experts for clarity, comprehensiveness, and ease of use. Four Community Medicine experts not involved in its development reviewed the tool for relevance and comprehensiveness. The structured assessment tools used for the six case topics are provided in Appendices A-F.

Assessment procedure

Four faculty members from the Department of Community Medicine served as examiners (Examiners A, B, C, and D). All four received orientation to the structured assessment tool before the assessment. Each student underwent only one clinico-social case discussion and was assessed simultaneously by two examiners.

For the first two batches, the focus was on the comparison of assessment methods. Each student was simultaneously assessed by two examiners, with one examiner using the conventional assessment method and the other using the structured assessment tool. Examiner allocation was rotated across batches to minimize examiner-related bias.

For the next three batches (3-5), the focus was on the assessment of inter-rater reliability. Each student was assessed simultaneously by two examiners using the same assessment method. Students were alternatively allocated to either group according to their order of presentation. One group was assessed by two examiners using the conventional method, while the other group was assessed by two examiners using the structured tool. Examiner roles were alternated between batches. All examiners scored independently and were unaware of the scores assigned by the other examiner.

Supplementary faculty feedback

Written feedback was obtained from six faculty members with more than five years of teaching experience in Community Medicine using an open-ended questionnaire. Three faculty members had participated as examiners during the study, while three additional faculty members were oriented to and reviewed the structured assessment tool.

This process has been described graphically in Figure 1.

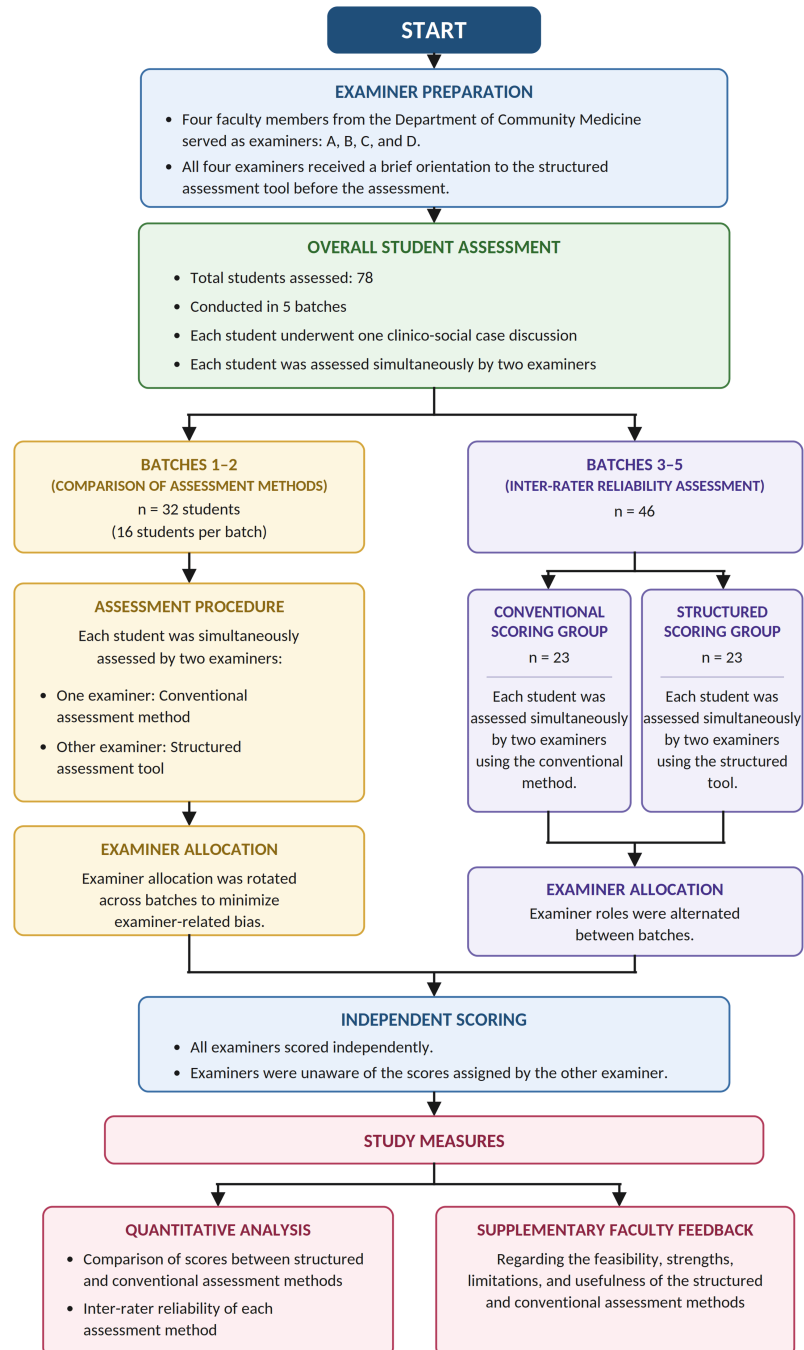


FIGURE 1: Flow diagram of study design and participant allocation

Data were analysed using IBM SPSS Statistics for Windows, Version 23 (Released 2016; IBM Corp., Armonk, New York, United States). Normality was assessed using the Shapiro-Wilk test. Scores obtained using the structured assessment tool and conventional assessment methods were compared using a paired t-test for mean score comparison, Pearson correlation, and Bland-Altman for agreement. Inter-rater reliability of structured and conventional assessment methods was assessed separately using the intraclass correlation coefficient (ICC). A two-way mixed-effects model with an absolute-agreement definition and single-

measurement type was used, with students treated as random subjects and examiners as fixed raters. Absolute agreement was selected to assess agreement between examiner scores, while the single-measure ICC reflected the reliability of an individual examiner's rating [12].

The written responses from faculty were reviewed descriptively and grouped under three pre-defined domains: experience with conventional assessment, feedback on structured assessment, and feasibility and implementation. Similar responses within each domain were summarized into major themes.

Approval was obtained from the Institutional Ethics Committee, Government Medical College Manjeri (IEC/GMCM/218/2026 dated 18/04/2026). Written informed consent was obtained from all participating students and faculty members before enrolment in the study. Participation in the study and scores obtained during assessment did not influence the academic assessment of students. Confidentiality of collected data was maintained throughout the study.

Results

A total of 78 students from five consecutive posting batches were included; each student was assessed by two examiners. Of the 78 participants, 53 (68.7%) were female, and the mean age was 21.1 (SD=1.56) years. Of these, 32 students from the first two batches were assessed simultaneously by two examiners using two different assessment methods. The remaining 46 students from the subsequent three batches were assessed by two examiners simultaneously using the same tool, with 23 assessed using the conventional assessment and 23 using structured assessment. All students had completed their Phase III Community Medicine clinical posting.

Distribution of scores (all batches)

Across all 78 students, the two assessment systems showed similar variability, with coefficients of variation of 14.5% for conventional and 14.0% for structured assessment.

Comparison of methods (batches 1 and 2)

Thirty-two students were assessed for the same presentation by two examiners using two different assessment tools. The paired differences between conventional and structured assessment scores did not significantly deviate from normality (Shapiro-Wilk $W=0.947$, $p=0.120$). Conventional assessment gave a significantly higher score than the structured assessment, with a mean difference of 1.31 points (SE=0.283, $t=4.64$; $df=31.0$; $P<0.001$). The scores showed a moderate positive correlation (Pearson's correlation $r=0.605$; $P<0.001$) (Figure 2).

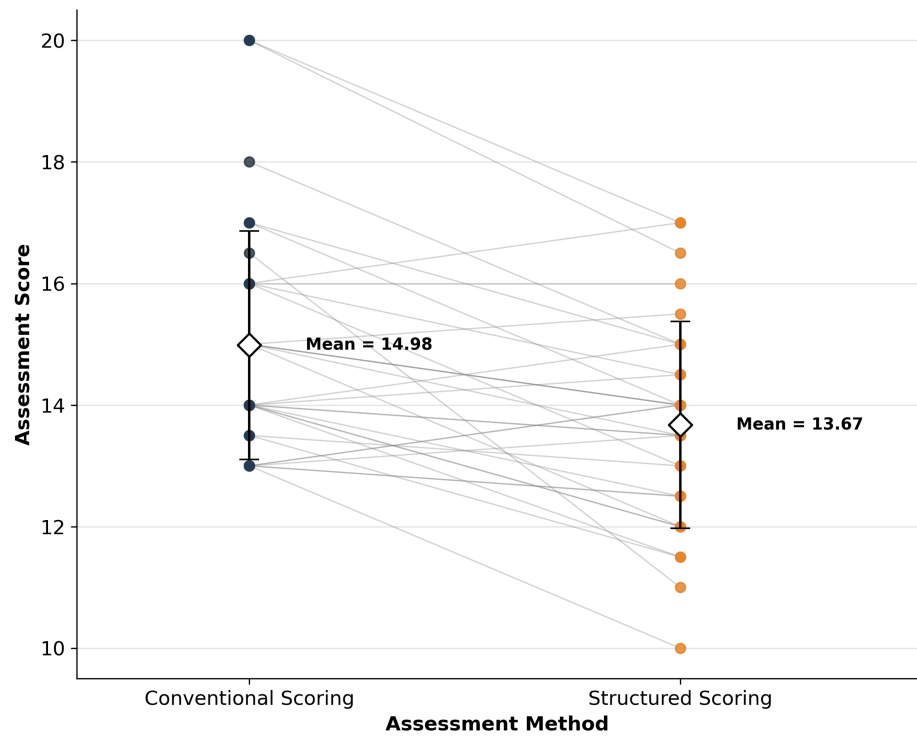


FIGURE 2: Comparison of individual student scores across conventional and structured assessment methods (n=32)

Each line connects the scores of the same student under the two assessment methods. Diamonds represent the group mean assessment score, and error bars represent ± 1 standard deviation (SD).

Agreement between methods

Bland-Altman analysis was performed to assess agreement between the two assessment methods and visualize systematic bias across the score range. The mean bias (conventional minus structured score) was 1.312 (95% CI 0.736 to 1.889), indicating a systematic tendency for structured scores to be lower than conventional scores. The standard deviation of the differences was 1.6 points, with limits of agreement ranging from -1.824 to 4.449. One of 32 observations (3.1%) fell outside the limits of agreement (Figure 3).

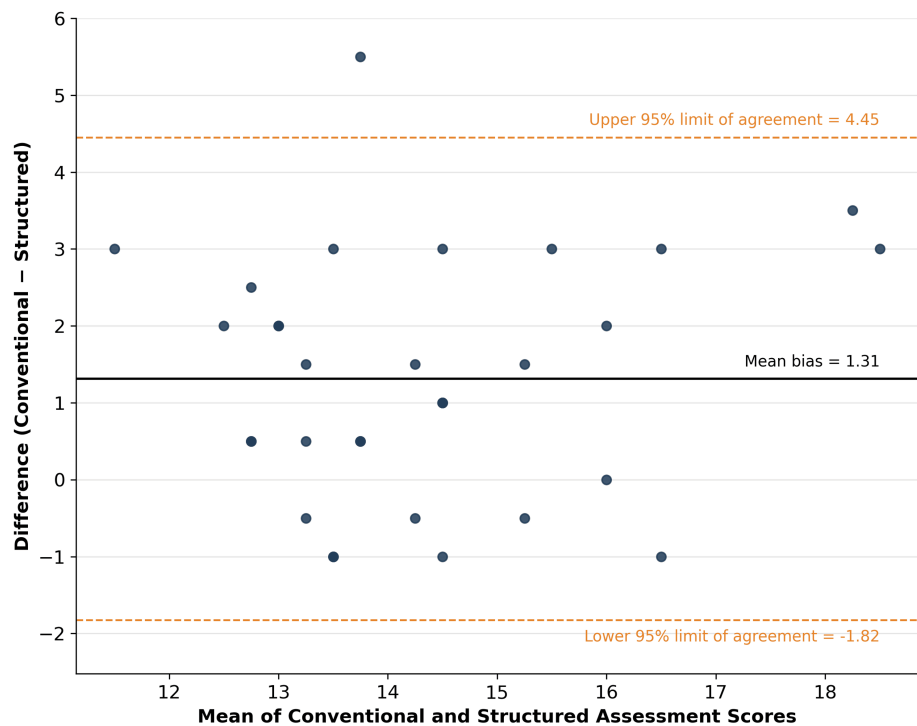


FIGURE 3: Bland-Altman analysis of agreement between conventional and structured assessment scores

Inter-rater reliability (batches 3-5)

Forty-six students were assessed simultaneously by two examiners using the same method to assess inter-rater reliability: 23 by conventional assessment and 23 by structured assessment. The score distributions in the two assessment groups also showed no significant departure from normality (all $p > 0.05$). Both methods showed good inter-rater reliability, with a higher point estimate for structured assessment (ICC 0.826; 95% CI 0.636-0.922) than for the conventional tool (ICC 0.782; 95% CI 0.577-0.901). However, the confidence intervals overlap substantially (Figure 4).

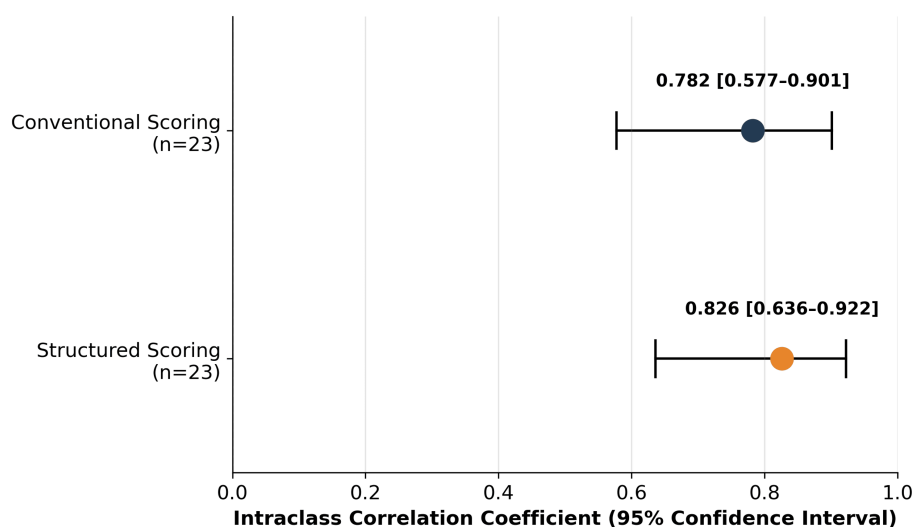


FIGURE 4: Inter-rater reliability of conventional and structured assessment methods

Intraclass correlation coefficients were calculated separately for the conventional and structured scoring groups (n=23 students per group, rated by two examiners) using a two-way mixed-effects model with absolute agreement and single-measurement type. Error bars represent 95% confidence intervals.

Faculty feedback

Conventional Assessment

In their feedback, faculty members identified flexibility as the major strength of the conventional clinico-social case discussion assessment. The open-ended nature of the assessment allowed examiners to individualize discussions based on the student's responses and explore issues relevant to each case, especially in complex clinical and social scenarios where patient context and determinants of health vary widely. However, the absence of predefined scoring criteria was considered an important limitation. Major concerns included subjectivity, examiner-dependent variation, differences in areas emphasized during discussion, and difficulty in converting overall performance into a numerical score.

Structured Assessment

Faculty feedback indicated that a structured assessment format could improve clarity, standardization, and coverage of essential competencies. They considered that predefined domains and scoring criteria help reduce examiner variability, ensure uniform assessment of core areas, and provide a more transparent scoring framework. It could also guide student preparation with structured feedback. Major concerns were excessive structuring, including reduced flexibility, increased time requirements, and difficulty in applying a fixed format to diverse clinico-social case scenarios.

Feasibility of implementation

A structured framework with clearly defined core domains, while retaining flexibility for case-specific exploration, was considered more practical than a rigid checklist-based approach. Suggested measures for improving assessment quality included examiner training, standardization of scoring criteria, adequate time allocation, and multiple opportunities for assessment. The main themes from the faculty feedback are summarized in Table 1.

Domain	Major theme	Key points
Conventional assessment	Flexible but examiner-dependent	Allows holistic case discussion and exploration of social context; limited by subjectivity and variation in scoring
Structured assessment	Improves standardization and clarity	Ensures competency coverage, transparent scoring, and consistency between examiners
Challenges of structured assessment	Need to avoid excessive rigidity	Concerns regarding reduced flexibility, increased time requirement, and applicability across different cases
Implementation	Structured framework with flexibility	Faculty training, defined core domains, simple scoring system, and multiple assessments suggested

TABLE 1: Summary of faculty feedback on clinico-social case assessment methods

Discussion

In this study, we compared a structured assessment tool with the conventional method for clinico-social case discussion in Community Medicine. The structured assessment tool and conventional method demonstrated comparable overall variability in scores, but differences were observed in scoring patterns. Structured assessment scores were systematically lower than conventional scores, with a mean difference of approximately 1.3 points. Bland-Altman analysis demonstrated a relatively consistent bias across the score range, while the moderate positive correlation indicated that students receiving higher scores with one method generally tended to receive higher scores with the other method. The lower scores observed with the structured tool may reflect differences between criterion-based assessment, where marks are assigned across predefined domains, and conventional global assessment, where examiners provide an overall judgement of student performance. The structured tool may have also made the domain-specific deficiencies more explicit, potentially contributing to the observed difference in scores.

The findings are consistent with the broader literature on long case assessment, which recognizes both the educational value and psychometric challenges of traditional case-based examinations. Ponnampuruma et al., in their review of the long case and its modifications, highlighted that the long case provides authentic assessment of clinical competence but suffers from limitations related to reliability due to inadequate sampling, case specificity, and examiner variability [13]. Singh has similarly emphasized that the strength of the long case lies in assessing integrated clinical performance, and modifications should aim to improve its quality without losing the holistic nature of the encounter [14].

Both methods demonstrated good inter-rater reliability, with a higher point estimate observed for the structured assessment method (ICC 0.826 versus 0.782). The number of students in each reliability subgroup limits the precision of these estimates. Examiner familiarity with the conventional method may have contributed to the observed difference in inter-rater reliability. Criterion-based scoring requires shared understanding of performance levels among assessors. Previous studies on OSLER have highlighted that introducing structure alone may not improve reliability unless accompanied by adequate examiner training and calibration [9,10].

The findings also highlight that standardization, objectivity, and reliability are related but distinct attributes of assessment. Excessive structuring or checklist-based approaches may improve standardization but may not necessarily enhance the assessment of complex competencies requiring professional judgement [6].

Faculty feedback provided additional context for interpreting these findings. Although examiners valued the flexibility of conventional assessment, they highlighted concerns regarding subjectivity and lack of explicit criteria. Conversely, the structured tool provided clearer domains and a transparent scoring framework but introduced challenges related to time, flexibility, and adaptation to diverse clinico-social scenarios. These findings align with previous work by Gohel et al., who demonstrated the feasibility of applying structured assessment principles to clinico-social case assessment in Community Medicine while maintaining the need for contextual judgement [11].

The educational implication of these findings is that structured assessment tools for clinico-social case discussions may be most useful when implemented as frameworks to support examiner judgement rather than as rigid checklists. Incorporating defined assessment domains, adequate examiner orientation, and repeated use may help improve consistency while maintaining the flexibility required for complex community-based case discussions. This study contributes to the limited literature on structured assessment of undergraduate clinico-social case discussions in Community Medicine by providing a direct comparison with conventional assessment methods.

The study has certain limitations. It was conducted in a single institution with a limited number of students and examiners, which may affect the generalizability of the findings. Although the inter-rater reliability estimates were reasonably high for both methods, the confidence intervals indicate some uncertainty in the precision of these estimates. The higher point estimate observed for the structured method should not be interpreted as evidence of superior reliability, and the findings require confirmation in larger, multicentre studies involving more examiners and multiple cases. The comparisons between methods should therefore be interpreted cautiously.

Assessment of complex competencies such as clinico-social reasoning is influenced by multiple sources of variability, including case specificity, student performance, examiner judgement, and assessment context. Increasing the number of cases, assessment occasions, and sampling across competencies may contribute more to improving reliability than focusing solely on reducing inter-rater variability [15]. The study design also introduced some potential sources of bias. Although examiner roles were alternated across batches and examiners scored independently, simultaneous assessment by two examiners does not fully replicate the usual examination setting in which a student is assessed by one examiner. Interaction between the student and one examiner may have influenced the assessment by the other examiner despite independent scoring. Examiner-related effects also cannot be eliminated, and the familiarity of the examiners with the conventional method may have influenced the scoring. Furthermore, clinico-social cases are inherently heterogeneous, and differences in case complexity, presentation, and scope for discussion may have influenced the applicability of the structured assessment and the resulting scores. Student perceptions were not assessed because students were not informed that different assessment methods were being used during the simultaneous assessment. Future qualitative studies could explore students’ perceptions of fairness, transparency, feedback, and the overall assessment experience.

Conclusions

The structured assessment tool provided a transparent framework with predefined assessment domains, expected elements, and scoring allocations for clinico-social case discussions and showed a higher point estimate of inter-rater reliability than the conventional method. Both methods demonstrated relatively high inter-rater reliability, while faculty feedback indicated that predefined criteria could improve clarity and consistency but should be implemented with sufficient flexibility for case-specific discussion. Further evaluation involving examiner calibration, multiple cases, larger samples, and multiple institutions is warranted before wider implementation.

Appendices

Appendix A

Structured Assessment Tools for Clinico-Social Case Discussions

Domain	Expected elements	Max marks	Marks awarded
Case summary and clinical reasoning	Structured presentation, relevant history, key findings, justified diagnosis	6	
Investigations and interpretation	Appropriate investigations and correct interpretation	3	
Social and behavioural determinants	Lifestyle, socioeconomic, environmental, adherence, family support	6	
Levels of prevention	Primary, secondary, tertiary prevention and gaps	4	
Management and public health linkage	Individual care, family advice, national program linkage, follow-up	4	
Communication and counselling	Patient education, adherence, follow-up advice	2	
Total		25	

TABLE 2: Non-communicable disease (NCD)

Appendix B

Domain	Expected elements	Max marks	Marks awarded
Case summary and clinical reasoning	Structured presentation, relevant history, key findings, justified diagnosis	5	
Investigations and interpretation	Appropriate investigations and correct interpretation	4	
Social and behavioural determinants	Lifestyle, socio-economic, environmental factors	6	
Levels of prevention	Primary, secondary, tertiary prevention and gaps	4	
Management and public health linkage	Individual care, national program linkage, follow-up	4	
Communication and counselling	Patient education, family advice, follow-up advice	2	
Total		25	

TABLE 3: Communicable disease

Appendix C

Domain	Expected elements	Max marks	Marks awarded
Case summary and clinical reasoning	Structured presentation, relevant history, key findings, justified diagnosis	5	
Investigations and interpretation	Appropriate investigations and correct interpretation	4	
Social and behavioural determinants	Lifestyle, socioeconomic, environmental, adherence, family support	6	
Levels of prevention	Primary, secondary, tertiary prevention and gaps	4	
Management and public health linkage	Individual care, family advice, national program linkage, follow-up	4	
Communication and counselling	Patient education, adherence, follow-up advice	2	
Total		25	

TABLE 4: Geriatric case

Appendix D

Domain	Expected elements	Max marks	Marks awarded
1. Case presentation and child assessment	Structured presentation, relevant history (including antenatal and birth history), key findings, justified diagnosis (if applicable)	6	
2. Growth, nutrition and development assessment	Growth parameters, nutritional assessment, feeding practices, developmental milestones, developmental concerns	5	
3. Family, environmental and social determinants	Family structure, socioeconomic status, caregiver practices, housing, sanitation, environmental exposures, childcare practices, health-seeking behaviour	5	
4. Preventive child health	Immunization, growth monitoring, deworming, missed preventive opportunities (if applicable)	4	
5. Management and public health linkages	Management plan, referral (if required), follow-up RBSK, UIP, ICDS, other national programmes	3	
6. Parent/caregiver counselling	Feeding advice, home care, danger signs, immunization advice, follow-up instructions	2	
Total		25	

TABLE 5: Under five child

RBSK: Rashtriya Bal Swasthya Karyakram; UIP: Universal Immunization Programme; ICDS: Integrated Child Development Services

Appendix E

Domain	Expected elements	Max marks	Marks awarded
Case summary and clinical reasoning	Structured presentation, relevant history, key findings, justified diagnosis	5	
Investigations and management	Investigations, routine PN care, special case if needed	6	
Antenatal and intranatal aspects	Details of antenatal and intranatal care	4	
Physiological aspects	Physiology of normal pregnancy and lactation	4	
Public health linkage	National program linkage, aspects covered in national program	4	
Communication and counselling	Dietary advice, family planning advice, follow-up advice	2	
Total		25	

TABLE 6: Postnatal case/lactating mother

PN: postnatal

Appendix F

Domain	Expected elements	Max marks	Marks awarded
Case summary and clinical reasoning	Structured presentation, relevant history, key findings, justified diagnosis	5	
Investigations and interpretation	Appropriate investigations and correct interpretation	4	
Social and behavioural determinants	Lifestyle, socioeconomic, environmental, adherence, family support	6	
Levels of prevention	Primary, secondary, tertiary prevention and gaps	4	
Management and public health linkage	Individual care, family advice, national program linkage, follow-up	4	
Communication and counselling	Patient education, adherence, follow-up advice	2	
Total		25	

TABLE 7: Antenatal case

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Jenyz M. Mundodan, Minu Mohan

Acquisition, analysis, or interpretation of data: Jenyz M. Mundodan, S. S. Roshni, Anupama Arumadi

Drafting of the manuscript: Jenyz M. Mundodan, S. S. Roshni

Critical review of the manuscript for important intellectual content: Jenyz M. Mundodan, S. S. Roshni, Anupama Arumadi, Minu Mohan

Disclosures

Human subjects: Informed consent for treatment and open access publication was obtained or waived by all participants in this study. Institutional Ethics Committee, Government Medical College Manjeri issued approval IEC/GMCM/218/2026. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors acknowledge the contributions of Dr. Priya Chandran, Professor and Head, and Dr. Biju George, Associate Professor, Department of Community Medicine, Government Medical College, Kozhikode, for their input in finalizing the study protocol. The authors also acknowledge the faculty members of the Department of Community Medicine for their contributions to the expert review and refinement of the structured tool, assessment as part of data collection, and faculty feedback. We also thank the students who participated in this study. Claude AI (Anthropic, San Francisco, California, USA) was used to assist in preparing the study flow diagram and in generating Python code used to create the spaghetti plot, Bland-Altman plot, and intraclass correlation coefficient (ICC) comparison plot. The final figures were generated from the resulting code using numerical data provided by the authors. The authors independently verified all data, calculations, labels, and graphical content. AI tools were not used for data collection, statistical analysis, interpretation of results, or generation of scientific conclusions.

References

1. Medical Council of India: Module for Undergraduate Medical Education. Medical Council of India, New

- Delhi, India; 2019.
2. van der Vleuten CP, Schuwirth LW: Assessing professional competence: from methods to programmes . Med Educ. 2005, 39:309-17. [10.1111/j.1365-2929.2005.02094.x](#)
 3. Norcini J, Anderson B, Bollela V, et al.: Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011, 33:206-14. [10.3109/0142159X.2011.551559](#)
 4. Prabhu SG, Abraham G, Nair LM: Comparative study of conventional assessment and OSLER conducted at the model clinical examination in a teaching hospital. Surg Rev Int J Surg Trauma Orthoped. 2020, 6:154-8. [10.17511/ijoso.2020.i03.03](#)
 5. Brannick MT, Erol-Korkmaz HT, Prewett M: A systematic review of the reliability of objective structured clinical examination scores. Med Educ. 2011, 45:1181-9. [10.1111/j.1365-2923.2011.04075.x](#)
 6. Singh T, Anshu: Is objectivity synonymous with reliability? . Principles of Assessment in Medical Education. Singh T, Anshu (ed): Jaypee Brothers Medical Publishers, New Delhi, India; 2024. 352-63. [10.5005/jp/books/18604_27](#)
 7. Norcini JJ, Blank LL, Duffy FD, Fortna GS: The Mini-CEX: a method for assessing clinical skills . Ann Intern Med. 2003, 138:476-81. [10.7326/0003-4819-138-6-200303180-00012](#)
 8. Gleeson F: AMEE Medical Education Guide No. 9: assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). Med Teach. 1997, 19:7-14. [10.3109/01421599709019339](#)
 9. Shah SI, Baig M, Shah S, Bashir EA, Sarwar H, Shah JA: Inter-rater reliability of Objective Structured Long Examination Record. J Ayub Med Coll Abbottabad. 2018, 30:180-3.
 10. Suresh RM, Venugopal K, Bindu CB, Halesha BR, Nachiketh NS: A comparative study of formative assessment in medicine using conventional long case examination and Objective Structured Long Examination Record (OSLER) among final-year MBBS students. Acad J Med. 2019, 2:151-4. [10.48165/rzh9dv46](#)
 11. Gohel M, Singh US, Bhandari D, Phatak A: Developing and pilot testing of a tool for "clinicosocial case study" assessment of community medicine residents. Educ Health (Abingdon). 2016, 29:68-74. [10.4103/1357-6283.188684](#)
 12. Koo TK, Li MY: A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016, 15:155-63. [10.1016/j.jcm.2016.02.012](#)
 13. Ponnampertuma GG, Karunathilake IM, McAleer S, Davis MH: The long case and its modifications: a literature review. Med Educ. 2009, 43:936-41. [10.1111/j.1365-2923.2009.03448.x](#)
 14. Singh T: The long case . Principles of Assessment in Medical Education. Singh T, Anshu (ed): Jaypee Brothers Medical Publishers, New Delhi, India; 2024. 83-90. [10.5005/jp/books/18604_8](#)
 15. Downing SM: Validity: on meaningful interpretation of assessment data . Med Educ. 2003, 37:830-7. [10.1046/j.1365-2923.2003.01594.x](#)