ORIGINAL ARTICLE
Journal of Science Technology Education Art and Medicine
DOI: 10.63137/jsteam.902684

Developing and Validating a Scoring Framework for the Assessment of Curriculum Viability in Undergraduate Medical Education

Fatima Sikander[a], Rehan Ahmed Khan[b], Saria Khalid[c], Saima Batool[d], Maria Shakeel[d], Khizar Ansar Malik[e*]

[a]Akhtar Saeed Medical and Dental College, Lahore, Pakistan
[b]Islamic International Medical College, Riphah International University, Islamabad, Pakistan
[c]University of Health Sciences, Lahore, Pakistan
[d]University College of Medicine and Dentistry, University of Lahore, Lahore, Pakistan
[e]Azra Naheed Medical and Dental College, Superior University, Lahore, Pakistan

*Corresponding address: *Azra Naheed Medical and Dental College, Superior University, Lahore, Pakistan*
Email: khizar.ansar@superior.edu.pk

Received: 06 April 2026 / **Revised:** 30 May 2026 / **Accepted:** 09 June 2026 / **Available Online:** 26 June 2026

ABSTRACT

Objectives: Curriculum viability indicators are among the factors that determine a curriculum's success. This study aims to develop and validate a scoring framework for assessing curriculum viability.
Methods: A cross-sectional quantitative study was conducted in two phases. The first phase involved constructing norm tables for interpreting the Curriculum Viability Student Questionnaire (CVSQ) and Curriculum Viability Teacher Questionnaire (CVTQ) using the median split method. The second phase focused on validating the proposed scoring framework through the Modified Angoff Method and Receiver Operating Characteristic (ROC) analysis. Participants included 281 undergraduate medical students and 93 faculty members from the University of Lahore, selected through convenience sampling.
Results: A total of 281 students and 93 faculty members participated. The CVTQ norm table established cut-off ranges of 25–50 (high viability), 51–75 (satisfactory), and 76–125 (low viability). ROC analysis identified an optimal CVTQ cut-off of 62.5 (sensitivity 84.6%, specificity 83.3%, AUC = 0.67, 95% CI: 0.531–0.716, p = 0.32); however, the AUC did not reach statistical significance. The Modified Angoff method established cut-offs of 42 for the CVTQ (ICC = 0.927) and 46 for the CVSQ (ICC = 0.897). ROC analysis was not feasible for the CVSQ because no gold-standard classification was available.
Conclusion: This study offers a comprehensive diagnostic tool and scoring framework for curriculum viability and a basis for informed decision-making in educational planning and improvement. The framework provides practical thresholds for identifying curricular strengths and problem areas, supporting evidence-based decisions in educational planning. Future research should test the scoring framework across diverse institutions to enhance its generalizability and explore additional approaches for validating student questionnaire cut-off scores.
Keywords: Curriculum; Educational measurement; Psychometrics; ROC curve; Education, Medical, Undergraduate; Surveys and questionnaires

Copyright: © 2026 by the authors. This is an open-access article licensed under the terms of the Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Sikander F, Khan RA, Khalid S, Batool S, Shakeel M, Malik KA. Developing and Validating a Scoring Framework for the Assessment of Curriculum Viability in Undergraduate Medical Education. J Sci Technol Educ Art Med. 2026;3(1):35-44

Introduction

A curriculum represents far more than a written document; it encompasses the totality of learning experiences designed, delivered, and evaluated within an academic program, including intended outcomes, content sequencing, teaching methods, assessment strategies, and the broader educational environment in which learning occurs. In medical education, where the stakes of curricular quality directly affect patient safety and clinical competence, the curriculum is widely regarded as a program's most valuable and dynamic asset, one that must evolve continuously to remain aligned with advances in science, pedagogy, and healthcare delivery.

Despite this recognition, curricular quality has historically been evaluated in a fragmented manner. Accreditation bodies such as the World Federation of Medical Education (WFME) and the Liaison Committee on Medical Education (LCME) set broad standards for educational programs.[1,2] However, their frameworks are primarily designed for periodic external review rather than for ongoing internal diagnosis. At the institutional level, curriculum evaluation has largely relied on tools that measure selected dimensions of the educational experience. The Dundee Ready Education Environment Measure (DREEM), for example, is one of the most widely used instruments globally.[3-5] However, it focuses specifically on students' perceptions of the educational environment and does not capture the structural, policy, or faculty-related inhibitors that may threaten curricular sustainability. Similarly, student satisfaction surveys and end-of-course evaluations provide useful but narrow snapshots that fail to address the multifactorial nature of curriculum health.

The concept of curriculum viability addresses this gap by shifting the focus from isolated dimensions of educational quality to a comprehensive assessment of whether a curriculum can sustain itself and fulfill its intended purpose over time. Khan et al.[6] defined curriculum viability as the current functional state of a curriculum, determined by the degree to which established benchmarks and standards have been met. Through a series of studies involving a scoping review, Delphi consensus, and instrument development, Khan and colleagues identified a set of curriculum viability indicators, factors that either support or inhibit the long-term health of a curriculum, and developed two validated instruments for their measurement: the Curriculum Viability Student Questionnaire (CVSQ) and the Curriculum Viability Teacher Questionnaire (CVTQ). The CVSQ comprises 14 items across three domains (educational program, student requirements, and institutional culture). In comparison, the CVTQ contains 25 items spanning six domains (educational program, disciplinary culture, social interaction, institutional policies, communication practices, and faculty involvement). Both instruments employ a five-point Likert scale and are designed to capture inhibitors, barriers that, when present, compromise the viability of the curriculum.[6-8]

The development and validation of the CVSQ and CVTQ represented a significant advance in curriculum evaluation methodology. However, a critical limitation remained: the instruments generate composite scores, but no empirically derived scoring framework exists to interpret them meaningfully. Without defined thresholds, an institution administering the CVSQ or CVTQ cannot determine whether a given score indicates a well-functioning curriculum, a curriculum requiring attention, or one at risk of failure. This interpretive gap limits the practical utility of the instruments. It prevents their use as diagnostic tools as clinical thresholds function in medicine, where, for example, a fasting blood glucose above 126 mg/dL or an HbA1c above 6.5% allows a clinician to classify a patient as diabetic and initiate appropriate management.

Establishing cut-off scores for educational measurement instruments is a well-recognized methodological need. Standard setting, the process of determining performance thresholds on a test or instrument, has been extensively studied in the context of licensing examinations and objective structured clinical examinations (OSCEs), where methods such as the Modified Angoff, Ebel, and borderline regression approaches are routinely applied.[9-12] However, the application of standard-setting methodology to curriculum evaluation instruments remains virtually unexplored. Similarly, Receiver Operating Characteristic (ROC) analysis, a widely used statistical technique in clinical medicine to determine optimal diagnostic thresholds, has seen limited application in educational research despite its potential to provide empirically derived cut-off scores with known sensitivity and specificity.[13]

This gap has practical consequences. Institutions that administer the CVSQ and CVTQ currently lack a framework for interpreting their results, assessing the health of their curricula, or making informed decisions about curricular reform. The absence of scoring benchmarks means the instruments remain descriptive rather than diagnostic, capable of generating data but unable to translate it into actionable classifications.

This study aimed to address this gap by developing and validating a scoring framework for the CVSQ and CVTQ that enables institutions to interpret scores, classify curricular viability, and identify specific domains requiring intervention. Specifically, the study sought to (1) establish empirically derived norm tables for interpreting composite and domain-level scores of the CVTQ and CVSQ, and (2) determine whether the CVTQ and CVSQ can function as diagnostic tools for curriculum viability, based on cut-off scores established through the Modified Angoff method and ROC analysis.

Materials and Methods

This study employed a cross-sectional, quantitative, two-phase design to develop and validate a scoring framework for curriculum viability indicators. Two reliable, pre-validated questionnaires, CVSQ and CVTQ, were used to establish the scoring framework. Both questionnaires employ a five-point Likert scale ranging from Strongly Agree to Disagree Strongly. The CVSQ comprises 14 items across three domains, whereas the CVTQ comprises 25 items across six domains.[8] Norm score validation was conducted using ROC analysis and the Modified Angoff method. The raw scores from the students' and teachers' responses were used as criteria for ROC analysis.

Phase I

In phase I, the median split method was used to determine the scoring framework as laid by DeCoster et al.[14] The original theoretical range of the CVSQ was used to establish cut-off boundaries rather than the empirically obtained score distribution from the study sample.[15] The cut-offs of the norm table were determined manually, with the median value of 3 serving as the midpoint. Tertile categorization was done to determine the cut-off range.[16] The tertile categorization is a method of dividing a data set into three equal parts, with two cut-off points (e.g., quartiles 1 and 3): one in the middle and two on the tails. Similar strategies for tertile categorization have been reported by other researchers as well.[16] Considering 3 as the median value on the 5-point Likert scale, the minimum and maximum possible scores for each domain and the overall instrument were calculated, and the resulting score ranges were categorized into tertiles as presented in the norm tables. Median splits score just above and just below the median 3 become high and low. So, a curriculum with weak inhibitors will have high viability, and vice versa. To use the CVTQ and CVSQ as a 'diagnostic tool' for a curriculum, each questionnaire subscale is intended to be used individually to diagnose problems in a curriculum, as well as the full CVSQ and CVTQ instruments.[4] The subscales of the CVTQ and CVSQ were constructed by summing the responses to the items that comprise each subscale.[3] Scores were added to calculate the overall scores and the subscale scores.[5]

In the curriculum viability teacher questionnaire, there were 25 items, and the CVTQ overall scores were set at 25-50 for high viability, 51-75 for satisfactory, and 76-125 for poor viability, as shown in Table 1.

After ethical approval from the ERB of the university (Ref no: ERC/126/23/12), written consent was obtained from all participants, including 281 students (2nd to final year MBBS) and 93 faculty members from The University of Lahore, each with at least two years of experience in curriculum planning and implementation. The data was collected through Google Forms, which were distributed to participants through the institutional communication channel, i.e., WhatsApp.

ORIGINAL ARTICLE
Journal of Science Technology Education Art and Medicine
DOI: 10.63137/jsteam.902684

Table 1: Cut-Off Scores of Curriculum Viability Teacher Questionnaire

Domain	No. of Items	Scores	Interpretations
Educational Program	6	6- 11	Weak Inhibitor
		12-18	Inhibitor neither present nor absent
		19-30	Strong inhibitor
Disciplinary Culture	2	2- 4	Weak Inhibitor
		5-6	Inhibitor neither present nor absent
		7-10	Strong inhibitor
Social Interaction	3	3-6	Weak Inhibitor
		7-9	Inhibitor neither present nor absent
		10-15	Strong inhibitor
Institutional Policies	5	5- 10	Weak Inhibitor
		11-19	Inhibitor neither present nor absent
		20-25	Strong inhibitor
Communication Practices	6	6- 11	Weak Inhibitor
		12-18	Inhibitor neither present nor absent
		19-30	Strong inhibitor
Faculty Involvement	3	3-6	Weak Inhibitor
		7-9	Inhibitor neither present nor absent
		10-15	Strong inhibitor
Overall	25	25- 50	High Viability
		51- 75	Satisfactory/ Optimum
		76-125	Low Viability

† CVTQ = Curriculum Viability Teacher Questionnaire. Scores based on a 5-point Likert scale (1 = Strongly Agree to 5 = Strongly Disagree). Cut-offs determined using the median split method with tertile categorization. Higher composite scores indicate a greater presence of curriculum inhibitors (lower viability). In the curriculum viability student questionnaire, there were 14 items, and the CVSQ overall scores were set at 14-28 for high viability, 29-42 for satisfactory, and 43-70 for poor viability, as shown in Table 2. This underscores the importance of scoring each domain, and the table makes it easy to interpret the scores for each inhibitor.

Table 2: Cut-Off Scores of Curriculum Viability Student Questionnaire

Domain	No. of Items	Scores	Interpretations
Educational Program	5	5- 10	Weak Inhibitor
		11-19	Inhibitor neither present nor absent
		20-25	Strong inhibitor
Students Requirements	5	5- 10	Weak Inhibitor
		11-19	Inhibitor neither present nor absent
		20-25	Strong inhibitor
Institutional Culture	4	4-8	Weak Inhibitor
		9-15	Inhibitor neither present nor absent
		16-20	Strong inhibitor
Overall	14	14 -28	High Viability
		29- 42	Satisfactory
		43- 70	Low Viability

Phase II

After the norm table with cut-offs was developed, Receiver Operating Characteristics (ROC) analysis was done to determine the specificity, sensitivity, and predictive values for the desired cut-offs.

ROC analysis is vital to assess the diagnostic capability of a test. The objective was to test the CVTQ and CVSQ as a diagnostic tool for curriculum viability. To see if the diseased curriculum (non-viable) could be differentiated from a healthy(viable)one using these questionnaires. The Area under the ROC curve (AUC) gives important information about the interpretation of the diagnostic test. It ranges from 0 to 1. The higher the number, the better the test performance. To assess the ability of the CVSQ and CVTQ to discriminate between viable and non-viable curricula, ROC analysis was done.

The students' and faculty members' scores were converted to the classification scale using the scores listed in the norm table. The scores of each student and faculty, in the study group, were classified as "low viability" or "high viability" based on a norm table in the scale manual, with scores below certain values coded as "low viability"(0) and scores above as "high viability"(1), i-e (1-0 coding). In this analysis, data with a mean less than 3 were considered viable, and those with a mean of 3 or more were considered neutral or non-viable curriculum.[17]

Validity through the Modified Angoff Method

To test the scores of the norm table and strengthen the methodological rigor of this study, the Modified Angoff method was used. This is the most widely used and well-documented method in the literature because of its ease of use. The principal researcher acted as the moderator of the focus group meeting. Participants were invited via WhatsApp, and the time and venue were arranged based on their availability. The session began with a brief presentation by the principal investigator explaining curriculum viability, the Modified Angoff method, and the Extended Modified Angoff process. Five experts in medical education (SMES) were taken for this study.[9] Literature suggests using 5-30 experts for the Angoff method, depending on the level of stake (high or low).[9-12] At the beginning of the focus group discussion (FGD), a brief discussion to determine the borderline case was conducted. Consensus was first reached on defining a borderline curriculum. The panel defined a borderline curriculum as one that is minimally functional, meeting basic operational requirements but at the threshold of non-viability. Two rounds of focus group discussions were conducted to determine the cut-off scores for CVTQ and CVSQ. In round I, each judge independently rated each item based on a minimally functional curriculum. In round II, experts reviewed and discussed their ratings to reach a collective agreement. Final cut-off scores were computed as the average of all agreed ratings across items.

Results

The mean CVTQ composite score among faculty was 61.8 ± 11.4 (range: 39–92). Based on the norm table classification, 18 (19.4%) faculty responses fell within the high viability range (25–50), 61 (65.6%) within the satisfactory range (51–75), and 14 (15.1%) within the low viability range (76–125). The mean CVSQ composite score among students was 36.7 ± 8.5 (range: 16–61), with 96 (34.2%) classified as high viability (14–28), 149 (53.0%) as satisfactory (29–42), and 36 (12.8%) as low viability (43–70). All participant demographics are presented in Table 3.

Results of Phase II

For ROC analysis, the data were collected from students and faculty using the CVSQ and CVTQ, respectively. In Table 4, for the ROC analysis, when the cut-off score is set to 62.5, the corresponding sensitivity and specificity are 84.6% and 83.3%, respectively. This cut-off was identified using Youden's J index (sensitivity + specificity − 1 = 0.679), which was the highest among all candidate thresholds. The reason for determining this value as the cut-off score is that both sensitivity and specificity for the scale refer to the best point at the same time.

ORIGINAL ARTICLE
Journal of Science Technology Education Art and Medicine
DOI: 10.63137/jsteam.902684

Table 3: Demographic Characteristics of Participants

Student's Demographics	n (%)
Gender	
Male	120 (42.7)
Female	161 (57.3)
Year of study	
2nd year MBBS	60 (21.4)
3rd year MBBS	68 (24.2)
4th year MBBS	70 (24.9)
5th year MBBS	83 (29.5)
Faculty Demographics	
Basic	33(35.5)
Clinical	53(57.0)
Para-Clinical	7(7.5)
Designation	
Professor	11(11.8)
Associate Professor	15(16.1)
Assistant Professor	12(12.9)
PG Trainee/Lecturer/Medical Officer	48(51.6)
Senior Registrar/Senior Lecturer	7(7.5)
Experience	
Less than 5 years	48(51.6)
5-10 years	26(28.0)
11-15 years	6(6.5)
16-20 years	9(9.7)
More than 20 years	4(4.3)

† MBBS = Bachelor of Medicine, Bachelor of Surgery. Percentages may not sum to 100% due to rounding.

Table 4: Cut-Off Scores of CVTQ as Determined by ROC Analysis

Cut offs	Sensitivity	Specificity	Cut offs	Sensitivity	Specificity
47.0000	1.000	1.000	72.5000	.641	.481
48.5000	1.000	.981	73.5000	.590	.444
49.5000	.949	.981	74.5000	.590	.389
50.5000	.923	.963	75.5000	.590	.370
52.0000	.897	.963	76.5000	.590	.315
53.5000	.872	.963	78.5000	.538	.278
54.5000	.872	.944	79.5000	.487	.259
56.0000	.872	.926	80.5000	.462	.259
58.0000	.872	.907	81.5000	.436	.259
59.5000	.872	.870	82.5000	.359	.222
61.0000	.872	.852	83.5000	.359	.204
62.5000	.846	.833	85.5000	.359	.185
63.5000	.821	.815	87.5000	.231	.148
64.5000	.821	.778	88.5000	.179	.130
65.5000	.769	.741	91.0000	.179	.093
66.5000	.744	.704	93.5000	.179	.074
67.5000	.744	.630	94.5000	.128	.074

68.5000	.744	.611	96.0000	.103	.056
69.5000	.718	.556	98.0000	.077	.056
70.5000	.692	.537	99.5000	.000	.019
71.5000	.641	.537	101.0000	.000	.000

† CVTQ = Curriculum Viability Teacher Questionnaire. Boldface indicates the selected optimal cut-off score (62.5). Sensitivity = proportion of correctly identified non-viable curricula; Specificity = proportion of correctly identified viable curricula. AUC = 0.67 (95% CI: 0.531–0.716, p = 0.32). Cut-off selected using Youden's J index (sensitivity + specificity − 1). Youden's J at 62.5 = 0.846 + 0.833 − 1 = 0.679.

The AUC was 0.67, which is acceptable according to the rule of thumb of Hosmer and Lemeshow in Applied Logistic Regression (p. 177),[13] with a p-value of 0.32 and a confidence interval of 0.531 (lower bound) and 0.716 (upper bound). The AUC of 0.67 (95% CI: 0.531–0.716, p = 0.32) indicated modest discriminatory ability that did not reach statistical significance. While the AUC exceeds 0.50 (chance level), the non-significant p-value suggests that the CVTQ's ability to discriminate between viable and non-viable curricula was not statistically demonstrated in this sample. This may reflect the limited sample size or the inherent difficulty of dichotomizing curriculum viability.

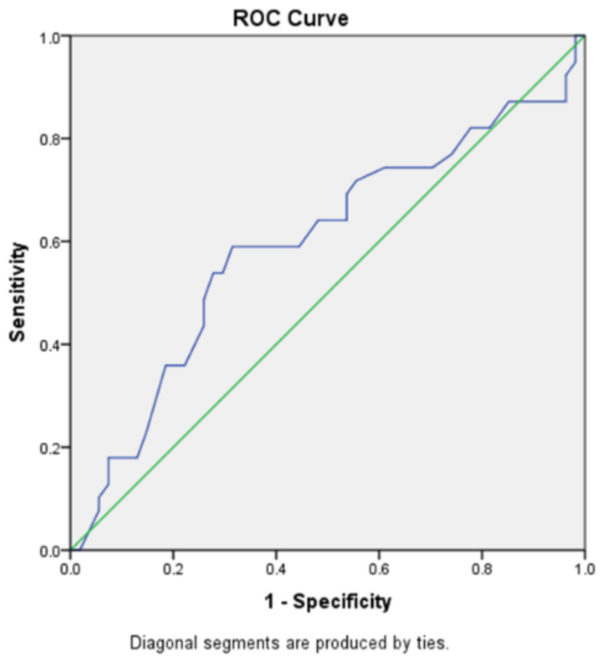

Figure 1: Receiver Operating Characteristic (ROC) curve for the Curriculum Viability Teacher Questionnaire (CVTQ). The diagonal reference line represents chance discrimination (AUC = 0.50). AUC = 0.67 (95% CI: 0.531–0.716, p = 0.32). The optimal cut-off score of 62.5 (sensitivity = 84.6%, specificity = 83.3%) is indicated.

The ROC curve showed that the CVTQ performed marginally above the no-discrimination line. However, the non-significant AUC (p = 0.32) indicates that this performance did not differ from chance in the current sample. The ROC-derived cut-off of 62.5 should therefore be interpreted cautiously and requires validation in larger, multi-institutional samples before being recommended for diagnostic use. In the Modified Angoff method, the cut-off of 42 was determined by the Subject experts after 2 rounds, with Cronbach's alpha and Interclass coefficient of 0.927 for the CVTQ. Whereas, the cut-off of 46 was determined by the subject experts after 2 rounds, with Cronbach's alpha and the interclass coefficient of 0.897 for the CVSQ.

Discussion

The curriculum serves as the cornerstone of any institute. Regular and systematic evaluation of curricula against defined standards is therefore vital to ensure their continued relevance, effectiveness, and alignment with institutional goals.[18] In this context, the Curriculum Viability Teacher and Student Questionnaire serves as a vital tool to diagnose the health status of a curriculum. To the best of our knowledge, existing tools were limited in scope, and the DREEM questionnaire, although widely used for evaluating the educational environment, addresses only part of the broader curriculum context.[5] In contrast, this study can help to determine curricular strengths and weaknesses by integrating perspectives from both students and teachers, the major stakeholders of a curriculum.

The present study provides a systematic scoring framework that enables institutions to assign quantifiable values to their curriculum, allowing decision-makers to distinguish between viable, borderline, and non-viable programs. This diagnostic approach supports informed educational planning and facilitates early identification of inhibitors within the system. The institutional culture, curricular structure, and student demographics at one specific university may have influenced the threshold values obtained. However, these potential limitations were minimized through expert validation using the Modified Angoff method. Two complementary methods were used to validate the scores. The Modified Angoff method, which uses expert consensus to capture the opinions and judgments of subject matter experts[16], is used to determine cut-off scores.

Meanwhile, the ROC analysis provided a statistical approach for determining the cut-offs for the CVSQ and CVTQ. The AUC of 0.67 falls below the acceptable threshold for discriminatory ability. Since the findings under study are not life-altering, a model with an AUC as low as 0.6 could still be useful. This approach established a quantitative link between psychometric validation and expert interpretation. Using these two methods reinforced the validity and reliability of the developed framework.

The three validation approaches yielded divergent CVTQ cut-off values: 51 (norm table lower boundary for satisfactory viability), 62.5 (ROC optimal cut-off), and 42 (Modified Angoff consensus). This divergence reflects the different methodological assumptions underlying each approach. The norm table boundaries are distributional, derived from the theoretical score range using median splits. The ROC cut-off maximizes the trade-off between sensitivity and specificity based on empirical data. The Modified Angoff cut-off reflects expert judgment of the score expected from a minimally functional curriculum. For practical application, it is recommended to use the norm table ranges (Table 1) for benchmarking and relative classification, and the Modified Angoff cut-off (42 for CVTQ, 46 for CVSQ) to identify curricula that fall below a minimum threshold of functionality. The ROC-derived cut-off requires further validation before clinical application.

The methodological approach adopted in this study parallels efforts in other fields to establish interpretive thresholds for self-report instruments. Yu et al.[15] used a combination of ROC analysis and norm-referenced methods to determine cut-off scores for the Family Burden Interview Schedule, similarly finding that different statistical approaches yielded divergent thresholds and recommending that the instrument's intended purpose guide the choice of method. Behlouli et al.[19] applied neural network-based methods to establish cut-off scores for the Minnesota Living with Heart Failure questionnaire, reinforcing the principle that self-report instruments require empirically derived scoring benchmarks before they can serve as diagnostic tools. However, both of these studies benefited from clinical gold standards against which sensitivity and specificity could be evaluated — an advantage not available for the CVSQ, where no external criterion for curriculum viability currently exists. The absence of such a gold standard is not unique to this study; it reflects a broader challenge in educational measurement, where constructs such as curriculum quality, institutional effectiveness, and teaching excellence lack universally accepted reference standards.

The Modified Angoff cut-off values obtained in this study (42 for CVTQ, 46 for CVSQ) demonstrated strong interrater reliability (ICC = 0.927 and 0.897, respectively), which compares favorably with Angoff-based standard-setting studies in clinical assessment contexts.[20-22] Shulruf et al.[20] reported interrater reliabilities of 0.85–0.92 when applying Modified Angoff procedures to OSCE stations, and Kamath et al.[21] similarly found acceptable consensus among expert panels when setting cut-offs for objective structured practical examinations. The relatively high ICC values in the present study may be attributable to the structured two-round deliberation process and the use of a clearly defined borderline construct — the minimally functional curriculum — which provided panelists with a concrete reference point for their ratings. Notably, however, the Angoff-derived cut-off of 42 for the CVTQ falls within the high viability range of the norm table (25–50), suggesting that expert judgment of minimal functionality sets a more lenient threshold than the distributional midpoint. This finding is consistent with the literature, which suggests that Angoff panels tend to set lower cut-offs than criterion- or norm-referenced methods, particularly when panelists are conservative in their estimates of the performance of borderline cases.

Five experts were recruited to rate the questionnaires, a number constrained by the availability of qualified subject matter experts at the study institution. Future studies can expand the panel size to explore whether increasing the number of SMEs changes the cut-offs. Another limitation of this study was that it was not possible to test the cut-off scores of the curriculum viability students' questionnaire due to the absence of a gold standard. Future researchers can look into alternative methods to validate the cut-off scores of the curriculum viability student questionnaire (CVSQ). Since the data were collected from a single institute, the generalizability of the results is limited.

The non-significant AUC ($p = 0.32$) represents a limitation of the ROC validation. In low-stakes educational assessment contexts, Hosmer and Lemeshow[13] suggest that an AUC between 0.6 and 0.7 may still provide acceptable discrimination. However, the wide confidence interval (0.531–0.716) and non-significant p-value indicate that the CVTQ's diagnostic utility was not conclusively demonstrated. The norm table and Modified Angoff cut-offs, which do not depend on statistical discrimination, therefore serve as the primary components of the proposed scoring framework. Replication with larger faculty samples across multiple institutions is needed to determine whether the CVTQ achieves statistically significant discriminatory ability.

Despite these limitations, this study contributes a novel, empirically grounded framework for curriculum evaluation. There are many self-administered questionnaires with different attributes and measuring different domains, such as the AIM and WHOQOL, with no cut-off scores. Future researchers can use the results of this study to develop scoring frameworks for these questionnaires. Additionally, the use of convenience sampling may have introduced selection bias, as voluntary participation could favor respondents with stronger opinions about their curriculum. The two-round Modified Angoff process, while consistent with established methodology, may benefit from additional rounds in future studies to further stabilize expert ratings. Future research should also explore the test-retest reliability of the norm table classifications by administering the CVSQ and CVTQ to the same cohort at multiple time points to assess the temporal stability of viability classifications.

Conclusion

In conclusion, this research fills a significant gap in the literature on curriculum evaluation. It presents a novel scoring system that could strengthen how institutions evaluate and forecast curricula's efficacy. Readers can use the norm tables with complete cut-off values to assess the viability of their curriculum. This study makes a meaningful contribution to the field of curriculum evaluation by integrating quantitative analysis with expert-based standard setting. The proposed scoring framework reinforces continuous quality improvement within institutions and serves as a versatile model across diverse academic environments.

Acknowledgments

The authors wish to thank the participating institutions and students for their time and cooperation in this study.

Author Contribution

FS conceived the idea and designed the study, supervised the overall project, and performed the final editing and critical revision of the manuscript. RAK and SK contributed to the study methodology, statistical framework, and critical review of the manuscript for important intellectual content. SM and MS conducted the initial literature review, organized the references, and assisted in manuscript preparation. KAM contributed to literature research and background review. All authors read and approved of the final manuscript and agree to be accountable for all aspects of the work.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The dataset generated during this study is available from the corresponding author upon reasonable request.

Conflict of Interest

The authors have no conflicts of interest to disclose.

References

1. World Federation for Medical Education. Basic medical education: WFME global standards for quality improvement. The 2020 revision. Ferney-Voltaire: World Federation for Medical Education; 2020.
2. Liaison Committee on Medical Education. Functions and structure of a medical school: standards for accreditation of medical education programs leading to the MD degree. 2025-26. Washington (DC): Liaison Committee on Medical Education; 2025.
3. Roff S, McAleer S, Harden RM, Al-Qahtani M, Ahmed AU, Deza H, et al. Development and validation of the Dundee Ready Education Environment Measure (DREEM). Med Teach. 1997;19(4):295-9. doi:10.3109/01421599709034208.
4. Swift L, Miles S, Leinster SJ. The analysis and reporting of the Dundee Ready Education Environment Measure (DREEM): some informed guidelines for evaluators. Creat Educ. 2013;4(5):340-7. doi:10.4236/ce.2013.45050.
5. Atapattu NSBM, Kumari KKEIU, Pushpakumara KKA, Mudalige SKK. Adoption of Dundee DREEM questionnaire to assess the educational environment of an agriculture degree programme. Trop Agric Res Ext. 2015;18(1):22-30. doi:10.4038/tare.v18i1.5321.
6. Khan RA, Spruijt A, Mahboob U, van Merrienboer JJG. Determining 'curriculum viability' through standards and inhibitors of curriculum quality: a scoping review. BMC Med Educ. 2019;19(1):336. doi:10.1186/s12909-019-1759-8.
7. Khan RA, Spruijt A, Mahboob U, Al Eraky M, van Merrienboer JJG. Curriculum viability indicators: a Delphi study to determine standards and inhibitors of a curriculum. Eval Health Prof. 2021;44(3):210-9. doi:10.1177/0163278720934164.
8. Khan RA, Spruijt A, Mahboob U, Al Eraky M, van Merrienboer JJG. Development and validation of teacher and student questionnaires measuring inhibitors of curriculum viability. BMC Med Educ. 2021;21(1):405. doi:10.1186/s12909-021-02843-0.
9. Hurtz GM, Auerbach MA. A meta-analysis of the effects of modifications to the Angoff method on cut-off scores and judgment consensus. Educ Psychol Meas. 2003;63(4):584-601. doi:10.1177/0013164403251284.
10. Kampa N, Wagner H, Köller O. The standard setting process: validating interpretations of stakeholders. Large Scale Assess Educ. 2019;7:3. doi:10.1186/s40536-019-0071-8.
11. MacDougall M, Stone G. Fortune-tellers or content specialists: challenging the standard setting paradigm in medical education programmes. J Contemp Med Educ. 2015;3(3):133-42. doi:10.5455/jcme.20151019104847.
12. Khalid MN, Shafiq F, Ahmed S. A comparison of standard setting methods for setting cut-scores for assessments with constructed response questions. Pak J Educ Res Eval. 2022;9(2):65-75.
13. Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2013. p. 177.
14. DeCoster J, Iselin AMR, Gallucci M. A conceptual and empirical examination of justifications for dichotomization. Psychol Methods. 2009;14(4):349-66. doi:10.1037/a0016956.
15. Yu Y, Liu ZW, Zhou W, Zhao M, Tang BW, Xiao SY. Determining a cutoff score for the Family Burden Interview Schedule using three statistical methods. BMC Med Res Methodol. 2019;19:93. doi:10.1186/s12874-019-0734-8.
16. Williams KA, Kanan S, Phillips L. Application of the Angoff method for assessing item appropriateness and setting cut-off scores during the pre-clinical instrument development phase. Madridge J Nurs. 2018;3(1):113-7. doi:10.18689/mjn-1000120.
17. Garcia D, MacDonald S, Archer T. Two different approaches to the affective profiles model: median splits (variable-oriented) and cluster analysis (person-oriented). PeerJ. 2015;3:e1380. doi:10.7717/peerj.1380.
18. DiFlorio I, Duncan P, Martin B, Middlemiss MA. Curriculum evaluation. Nurse Educ Today. 1989;9(6):402-7. doi:10.1016/0260-6917(89)90095-6.
19. Behlouli H, Feldman DE, Ducharme A, Frenette M, Giannetti N, Grondin F, et al. Identifying relative cut-off scores with neural networks for interpretation of the Minnesota Living with Heart Failure Questionnaire. Conf Proc IEEE Eng Med Biol Soc. 2009;2009:6242-6. doi:10.1109/IEMBS.2009.5334659.
20. Shulruf B, Damodaran A, Jones P, Kennedy S, Mangos G, O'Sullivan AJ, et al. Enhancing the defensibility of examiners' marks in high stake OSCEs. BMC Med Educ. 2018;18(1):10. doi:10.1186/s12909-017-1112-z.
21. Kamath MG, Pallath V, Ramnarayan K, Kamath A, Torke S, Gonsalves J. The standard setting of objective structured practical examination by modified Angoff method: a pilot study. Natl Med J India. 2016;29(3):160-2.
22. Thinggaard E, Bjerrum F, Strandbygaard J, Gögenur I, Konge L. Ensuring competency of novice laparoscopic surgeons-exploring standard setting methods and their consequences. J Surg Educ. 2016;73(6):986-91. doi:10.1016/j.jsurg.2016.05.008.