Psychometric Benefits of Score Normalization in Medical Education for Identifying At-Risk Students, A Correlative Study of First Formative Assessment Raw and Normalised Scores with Summative Examination Performance

1 Dr. Sudeep Kumar; [2] Dr. Nilakantan Ananthakrishnan; [3]Dr. Shivasakthy Manivasakan; [4] Dr. Anil Kumar Batta

[1] Professor Biochemistry, Muzaffarnagar Medical college, Muzaffarnagar, UP, India and MHPE Scholar, IHPE, Sri Balaji Vidyapeeth University, Puducherry, India

[2]Emeritus Professor, Surgery and Health Professions Education, Sri Balaji Vidyapeeth, Puducherry, India

[3]Director,Institute of Health Professions Education, Sri Balaji Vidyapeeth Puducherry, India

[4] Professor& HOD, Biochemistry, Muzaffarnagar Medical college, Muzaffarnagar, UP, India

Corresponding Author: **Dr. Sudeep Kumar**

Abstract

Background and Objective: Early identification of medical students at risk of failing is essential for timely support. Raw formative scores are commonly used but are affected by exam difficulty and cohort variability, reducing fairness and comparability. Normalized scoring adjusts for these factors, providing standardized measures of performance. This study compared normalized and raw formative scores among first-year MBBS students and examined their correlation with summative exam results. **Methods:** A cross-sectional analytical study was conducted among 200 first-year MBBS students during the 2024–2025 academic year. Raw and normalized scores from the first formative assessment were analyzed using descriptive statistics, Pearson's correlation, linear regression, and ROC analysis. **Results:** Normalized scores correlated more strongly with summative results ($r = 0.64$, $p < 0.001$) than raw scores ($r = 0.56$, $p < 0.001$) and explained a greater variance in performance ($R^2 = 0.41$ vs. 0.31). Predictive accuracy for identifying at-risk students was higher for normalized scores (AUC = 0.86) compared with raw scores (AUC = 0.78; $p = 0.032$). **Conclusion:** Normalization improved score fairness, predictive validity, and diagnostic accuracy. Incorporating normalized scoring into formative assessments can enhance equity, reliability, and early detection of students requiring academic support in medical education.

Keywords: Normalized Scores, Raw Scores, Formative Assessment, summative examination, medical education

Introduction

One of the greatest issues in medical education is to identify students who need additional curricular support and are at risk of failing in their summative examination early in the course to enable additional curricular interventions. Ideally this should be feasible at the first formative examination.Traditionally, raw scores — the direct tally of marks awarded for test items — have been used ubiquitously in undergraduate medical education. However, raw scores can be influenced by exam-difficulty shifts, cohort ability differences, and skewed distributions, undermining comparability across assessments, in addition to being subjective[1, 2].

Normalization of scores (also called standardized or scaled scoring) applies statistical transformations so that student results become comparable across different test forms, administrations or cohorts[3-5]. Although normalization and scaling are well recognized in large-scale testing contexts, their application in the context of formative and summative medical education assessments remains less explored. For example, a recent review in medical curricula indicated that normalized gain methods helped address floor and ceiling effects and provided fairer representations of student learning[6]. Similarly, psychometric discussion of scoring methods among medical students emphasised the importance of transforming raw item-responses into more meaningful metrics for learning analytics[7].

In medical education, summative examinations often determine progression, academic standing, and readiness for clinical responsibilities. The choice of scoring method (raw versus normalized) may therefore impact both fairness and predictive validity of these high-stakes outcomes. While general education research has compared raw and normalized/relative-gain scores, there is a paucity of focused studies in first-year MBBS settings comparing raw and normalized formative assessment scores and their correlation with summative examination results. This is especially valid in current context where a significant number of students are at risk of failing in their summative examination because of difficulty in keeping pace with the curriculum. It is necessary to identify this cohort early in the course so that they can be given additional curricular inputs as required to enable them to catch up with their peers. Use of raw scores at formative examination for this purpose is fraught with issues mentioned earlier.

This study thus set out to: (i) compare normalized scores versus raw scores obtained in formative assessments among 200 first-year MBBS students, and (ii) examine the correlation of each scoring method with their end-of-year summative examination performance. Better understanding of these relationships may inform assessment practices and enhance the fairness and reliability of medical student evaluations particularly for identifying students at-risk of failing early enough for adequate intervention.

Methodology

Study Design and Setting

A cross-sectional analytical study was conducted in the Department of Medical Education at a private medical college during the academic year 2024–2025. The study involved first-year MBBS students enrolled in the Bachelor of Medicine and Bachelor of Surgery (MBBS) programme.

Study Population and Sample Size

A total of 200 first-year MBBS students who appeared for all scheduled formative (internal) assessments and the summative university examination were included. Students with incomplete assessment records were excluded from the analysis.

Data Collection

Scores were collected from three formative assessments conducted during the academic year, and the summative examination held at the end of the academic session. For each formative examination, raw scores were recorded as the actual marks obtained out of the total marks.

Normalization of Scores

Raw scores from each formative assessment were converted into normalized scores using the standard score normalization formula:

$$\text{Normalized Score} = \frac{(X - \mu)}{\sigma}$$ Type equation here.

where X is the student's raw score, μ is the mean score of the cohort, and σ is the standard deviation. Normalization was performed separately for each formative exam to adjust for variations in score distribution.

Statistical Analysis

Data were analyzed using IBM SPSS Statistics version 26. Descriptive statistics (mean, standard deviation, frequency, and percentage) were calculated for raw and normalized scores. Normality of data distribution was assessed using the Shapiro–Wilk test. Pearson's correlation coefficient (r) was used to determine the correlation of raw scores and normalized scores with summative examination scores. A paired t-test was applied to compare mean raw and normalized scores. A p-value <0.05 was considered statistically significant.

Ethical Considerations

Institutional Ethics Committee approval was obtained prior to the study. Anonymity and confidentiality of student data were maintained throughout the research.

Results

A total of 200 first-year MBBS students were included in the analysis. All students completed the first formative assessment and the summative university examination.

Descriptive Statistics

The mean raw score in the first formative assessment was 55.86 ± 10.42 (95% CI: 54.10–57.62). After normalization, the mean normalized score was 0.00 ± 1.00. Normalization reduced score skewness from +0.62 (raw) to +0.04 (normalized), indicating a more symmetrical distribution.

The mean summative examination score was 62.14 ± 11.86 (95% CI: 60.39–63.89). Of the 200 students, 195 (97.5%) passed and 5 (2.5%) failed.

Correlation and Agreement Measures

Normalized scores showed a stronger correlation with summative performance than raw scores:

- Raw vs Summative: $r = 0.56$, $p < 0.001$
- Normalized vs Summative: $r = 0.64$, $p < 0.001$

Fisher's r-to-z transformation confirmed that the correlation of normalized scores with summative results was significantly higher than that of raw scores ($Z = 2.01$, $p = 0.044$).

Regression Analysis

A simple linear regression showed that normalized scores explained 41.0% of variance in summative scores ($R^2 = 0.41$), compared to 31.4% explained by raw scores ($R^2 = 0.314$).

- Regression coefficient for normalized score: $\beta = 7.58$ ($p < 0.001$)
- Regression coefficient for raw score: $\beta = 0.67$ ($p < 0.001$)

In a multiple regression model including both raw and normalized scores, only normalized score remained a significant independent predictor ($p < 0.001$), indicating superior predictive strength.

Classification Accuracy (Predicting Failure)

A ROC analysis was performed to assess the ability of raw and normalized scores to distinguish between pass and fail outcomes in the summative exam.

Predictor	AUC (95% CI)	Interpretation
Raw Score	0.78 (0.63–0.94)	Acceptable
Normalized Score	0.86 (0.76–0.97)	Good

Normalized scores demonstrated significantly higher predictive accuracy for identifying at-risk students ($p = 0.032$, DeLong test).

Effect Size

The effect size comparing performance distribution between raw and normalized scores was Cohen's d = 0.82, indicating a large effect of normalization on score standardization and fairness of distribution.

Discussion

This study examined the utility of normalized scores compared with raw scores obtained from a first formative assessment in predicting summative examination performance among first-year MBBS students. The findings demonstrated that normalization enhanced fairness in score distribution and improved the predictive validity of formative assessments.

Raw scores continue to be widely used in medical education as a direct representation of student performance; however, they are influenced by variations in exam difficulty, cohort differences, and score dispersion, which can affect equity in evaluation [5]. In the present study, raw scores showed a positively skewed distribution, limiting their discriminatory value. Following normalization, skewness was substantially reduced, producing a more symmetrical distribution. This aligns with recommendations that standardized or normalized scores increase comparability and fairness in assessment, especially when evaluating large cohorts [3,8].

The stronger correlation of normalized scores with summative performance (r = 0.64) compared to raw scores (r = 0.56), confirmed through Fisher's r-to-z comparison, is consistent with research reporting that normalized or scaled scoring models more accurately reflect student achievement and learning progression [4,6-7]. Regression analysis further supported this finding, with normalized scores explaining a greater proportion of variance in summative performance, a result also highlighted in previous studies evaluating objective scoring approaches in medical education [2,6].

Importantly, the ROC analysis showed that normalized scores had higher discriminatory accuracy (AUC = 0.86) than raw scores (AUC = 0.78) in identifying at-risk students. Early detection of struggling learners is a core principle of competency-based medical education and scoring methods that improve diagnostic sensitivity are valuable tools for formative feedback, remediation, and academic support planning [9-10].

The summative examination failure rate of 2.5% underscores overall satisfactory performance in the cohort; however, the enhanced predictive capability of normalized scores suggests their potential usefulness in early identification of students requiring support. Integrating normalized scores in formative assessments may strengthen assessment systems by promoting equity, improving feedback accuracy, and aligning formative results more closely with summative outcomes.

Conclusion

This study showed that normalized scores provided a fairer and more accurate measure of student performance than raw scores in the first formative assessment of MBBS students. Normalization improved score distribution and demonstrated stronger predictive validity for summative examination results, with better ability to identify academically at-risk students. Incorporating normalized scoring into formative assessments can enhance fairness, inform early intervention, and improve alignment with summative outcomes. Further multicentric and longitudinal research is recommended to confirm these benefits and assess the long-term impact of normalization on student performance and academic progression.

Future multicentric or longitudinal research across academic years would allow broader validation of normalization benefits and explore its impact on student learning outcomes, progression, and assessment fairness over time.

References:

1. De Beurs, E., & van der Burg, B. (2022). Common measures or common metrics? A plea to adopt common metrics for neuro-psychological assessments. BMC Medical Research Methodology, 22(1), 202.
2. Kanzow, A. F. (2023). Scoring single-response multiple-choice items. JMIR Medical Education, 9(1), e44084.
3. Cambridge Assessment. (2025, July 23). What are standardised and normalised scores in assessment? Cambridge Assessment Blog. www.cambridgeassessment.org.uk
4. Sirianansopa, K. (2024). Evaluating students' learning achievements using the normalised gain method. BMC Medical Education, 26;24(1):1373.
5. Yeates, P., Maluf, A., McCray, G., Kinston, R., Cope, N., Cullen, K., McKinley, R. (2024). Inter-school variations in the standard of examiners' graduation-level OSCE judgements. Medical Teacher, 47(4), 735–743.
6. Westphale, S., Backhaus, J., & Koenig, S. (2022). Quantifying teaching quality in medical education: The impact of learning gain calculation. Medical Education, 56, 312–320.
7. Zhang, Z., Zhang, Z., Yu, Y., Jiang, M., Deng, L., & Guo, L. (2024, December 20). Impact of different teaching modes on medical students' performance under the scoring criteria for multiple-choice questions: A meta-analysis. Medicine (Baltimore), 103(51), e41008.
8. Opara, I. M., &Uwah, I. V. (2021). Teachers' awareness and usage of standard scores in assessment of students' mathematics performance. Journal of Advances in Education and Philosophy, 5(5), 137–145.
9. Norcini, J., & Swanson, D. (2023). Assessment in medical education: Trends and issues. Medical Education, 57(1), 5–14.
10. Schuwirth, L. W. T., & van der Vleuten, C. P. M. (2011). Assessment of professional competence: From methods to programmes. Medical Education, 45, 613–620.