Nigerian Medical Journal
NIGERIA MEDICAL ASSOCIATION
Print ISSN 0300-1652, E-ISSN 2229-774X

Original Research

Disparities Impacts Psychiatry Competency Assessments among Students in a Nigerian Medical School

***Oluwabunnmi Ideraoluwa Nimata Buhari[1], Mumeen Olaitan Salihu[2], Adebusola Jane Ogunmodede[1], Peter Omoniyi Ajiboye[1].**

[1]Department of Behavioural Sciences, Faculty of Clinical Sciences, University of Ilorin/ University of Ilorin Teaching Hospital, Ilorin, Nigeria. [2]Department of Behavioural Sciences, Kwara State University, Teaching Hospital, Ilorin, Nigeria.

Abstract

Background: Assessments in medical education must reliably evaluate competencies while addressing potential gender disparities. This study examines the reliability of psychiatry competency assessments among final-year medical students at the University of Ilorin, Nigeria, and explores gender differences in performance. The aim was to evaluate the internal consistency of formative (progressive assessment) and summative (final MBBS) examinations, analyse correlations between assessment components, and compare gender-based performance.

Methodology: A retrospective analysis of all 137 (77 male, 60 female) final year medical students' psychiatry competence assessments was conducted. Reliability was measured using Cronbach's alpha, correlations between exam components were assessed via Pearson's correlation, and gender differences were analysed using t-tests and Cohen's d.

Results: The combined assessments showed high reliability ($\alpha = 0.857$), though lower for females ($\alpha = 0.553$ vs. males: $\alpha = 0.618$). Progressive assessment correlated moderately with theory components (Multiple Choice Questions: $r = 0.507$) but weakly with clinical tools (picture test: $r = 0.158$). Females outperformed males in both theory (mean difference: +1.71, $p = 0.001$) and clinical exams (mean difference: +1.08, $p = 0.019$), with moderate effect sizes (Cohen's $d = 0.63$ and 0.42, respectively).

Conclusion: While the assessment system demonstrates strong overall reliability, gendered disparities suggest a need for more equitable evaluation methods. Females consistently outperformed males, but lower reliability in their combined scores calls for refined formative tools. Integrating more multimodal assessments may enhance fairness and validity.

Keywords: Gender Disparities; Psychiatry Competency; Assessment; Formative; Summative; Evaluation; Medical Education.

***Correspondence:** Oluwabunmi Ideraoluwa Nimata, Buhari. Department of Behavioural Sciences, Faculty of Clinical Sciences, University of Ilorin/ University of Ilorin Teaching Hospital, Ilorin, Nigeria. **Email:** buhari.oin@unilorin.edu.ng.

How to cite: Buhari OIN, Salihu MO, Ogunmodede AJ, Ajiboye PO. Disparities Impacts Psychiatry Competency Assessments among Students in a Nigerian Medical School. Niger Med J 2025; 66 (4):1561-1571.https://doi.org/10.71480/nmj.v66i4.991.

Quick Response Code:

Introduction

Establishing clear expectations regarding the knowledge, skills, and attitudes required at various stages is fundamental to outcome-based medical education [1].Assessments serve as a vital mechanism for determining the extent to which medical students achieve the learning objectives that are outlined in each course [2].To ensure fairness and efficacy, assessments must align directly with these objectives and provide accurate evaluations of the students' performance [3].

Upon completing their undergraduate studies, medical graduates are expected to apply theoretical knowledge and demonstrate proficiency in clinical skills. The clinical competencies are typically evaluated through methods such as Objective Structured Clinical Examinations (OSCEs), Mini Clinical Evaluation Exercises (mini-CEXs), direct observation during clinical rotations, and case presentations, which assess students' understanding of clinical issues and their ability to analyze data [4].In contrast, cognitive abilities are often assessed through written examinations, case-based discussions, drills, and bedside teaching [4].

Notably, ensuring that students perform well in both cognitive examinations and clinical skills assessments presents several challenges to medical education [4,5].

These challenges include bridging the gap between theoretical knowledge and its practical application, addressing the inherent subjectivity of clinical evaluations, providing students with sufficient clinical exposure, and designing cognitive assessments that reflect the complexities of clinical practice, including adequate patient interaction and how to make decisions under pressure. Formative assessments, also known as progressive assessment (PA), highlight skill deficiencies that students may have, play a critical role in addressing these challenges, by offering structured remediation strategies [6].

Formative assessments are continuous, iterative evaluations designed to provide students with consistent feedback on their performance. They guide students in addressing knowledge gaps and enhance their conceptual understanding, ultimately improving their performance in summative assessments [7-9].

Summative assessments are the assessments conducted at the conclusion of a medical course or programme to evaluate a student's overall learning and competency. While summative assessments offer a comprehensive snapshot of a student's capabilities, they do not provide insights into the learning processes that led to the current level of competency or the potential trajectory for future development [8,10].

Bloom's Cognitive Learning Taxonomy, revised by David Kathwohl, provides a structured framework for designing assessments that ensure the progression of medical education from foundational knowledge to advanced competencies, such as diagnosing complex cases or developing innovative treatments [11,12]. The taxonomy describes the levels, in action forms, from lower levels to higher levels.

Written assessments in medical education focus on the lower levels of Bloom's Taxonomy (levels 1–3), which emphasize recalling, understanding, and applying information. Conversely, clinical examinations typically target the higher-order cognitive skills (levels 4–6), including application, analysis, and creation [13].

Furthermore, medical education necessitates the assessment of soft skills, such as teamwork and communication, to cultivate well-rounded and competent graduates. Achieving these objectives often requires the integration of multiple assessment tools to evaluate a wide array of learning outcomes effectively [14].

Numerous studies have tried to investigate factors influencing medical students' performance. For instance, personality qualities, level of motivation, curriculum design and psychological factors are some of the factors that have been found to influence performance among medical students [15-18]. Other studies identified gender as a contributing factor, where females were reported to often outperform male counterparts in academic and clinical assessments [19-23]. This success is attributed to their effective study strategies, higher academic motivation, better progress monitoring, and superior interpersonal and communication skills among females [19-23].

In the light of recent advancements in psychiatry and the evolving demands of healthcare, the University of Ilorin College of Health Sciences implemented a revision of its undergraduate Psychiatry medical education curriculum. Beginning with the 2023–2024 academic session, the program's duration was extended from six to eight weeks, and Psychiatry was restructured as an independent posting, distinct from Internal Medicine and the examination is now done as a separate course distinct from internal medicine. This final year posting builds upon foundational psychology lectures delivered during preclinical years.

The primary objective of this study was to assess reliability, gender differences, and correlations between examination components. It compared and analyzed the various components of medical students' psychiatry examinations, including the formative progressive assessment at the end-of-postings examination and the summative final MBBS examinations. It also evaluated gender differences and correlations between these examination components.

Materials and Methods

This study was a retrospective correlational study conducted at the Behavioural Sciences Department of the Faculty of Clinical Sciences, University of Ilorin, Ilorin, Nigeria. Based on the end-of-posting examinations (formative) and the final MBBS examination (summative), the performance of 137 medical students who were assigned to the Department for an eight-week program in the 2023–2024 academic session was analyzed.

The study included every final-year medical student in the 2023-2024 Academic session at the University of Ilorin who participated in both formative and summative assessments. Included were their gender, as well as their overall scores on the final MBBS assessments (comprising the written, clinical, oral, and picture tests), as well as their progressive assessment score.

Multiple choice and short answer essay (SAQ) questions made up the theory portion of the end-of-posting exam (progressive/ summative assessment), while the clinical component included long case presentations, picture tests, and portfolio evaluation. The final MBBS assessment includes multiple-choice questions (MCQs), SAQs, picture tests, OSCEs, and oral exams (viva). The progressive assessment score and final MBBS scores are then added to make the final total score to determine whether they are competent or not. Six stations make up the clinical examinations, comprising couplet stations for the OSCE, oral viva examination, procedure, and question stations.

This study involved a retrospective analysis and audit of anonymized student assessment data for the purpose of reviewing and improving assessment methods. The study utilized only de-identified data with no risk to participants, formal consent was not required but ethical approval was obtained (ERC PIN/2025/05/1571). The study complied with the Declaration of Helsinki's ethical guidelines, ensuring confidentiality and reducing risk to participants.

The data was coded and imported from Excel into SPSS version 24 for analysis. The data were then presented as percentages and frequencies. Cronbach's alpha was used to check for reliability. The consistency of students' performance across various examination components was assessed using Pearson's correlation (R). The gender differences across the progressive assessment and final MB examinations were assessed using t-test to compare means. The threshold for significance was set at a p-value of less than 0.05. Cohen's d (effect size) was used to assess the magnitude of differences between the means.

Results

In the 2023–2024 session cohort, a total of 137 students were evaluated, representing 100% of the class. Of these students, 77 (56.2%) were male and 60 (43.8%) were female.

Table 1 presents the internal consistency and reliability measures for both formative and summative assessments, in addition to other components. The aggregate result demonstrated high reliability (α = 0.857). Reliability was notably high for the final MBBS examination alone (α = 0.778). However, when combining progressive assessments (formative assessments) with the final examination, reliability was lower for females (α = 0.553). Conversely, males exhibited greater consistency and reliability in the combined results of theoretical knowledge and clinical performance (α = 0.729).

Table 1: Reliability scores of each component psychiatry examination

Variables	Cronbach's Alpha	Frequency
All items	0.857	137
Final MBBS alone (summative)	0.778	137
Progressive assessment(formative) & final MBBS (females alone)	0.553	60
Progressive assessment & final MBBS (males alone)	0.618	77
Final MBBS Theory and Clinical (females alone)	0.625	60
Final MBBS Theory and Clinical (males alone)	0.729	77

Table 2 presents the correlation between components (formative and summative assessments) of the psychiatry examinations. There were strong positive correlations (r > 0.5) between the progressive assessment (PA) and the MCQ and theory (SAQs + MCQs) of the final examination (0.507 vs 0.531). There was a moderate correlation between PA and MCQ (0.507), SAQ (0.322), OSCE (0.357), and Clinicals (0.315). A very weak correlation was observed with Oral (0.231) and a non-significant correlation with Picture test (0.158, p=0.066).

The theory component correlated strongly with MCQs (0.944) and moderately with SAQ (0.641). The Orals and Picture test had weaker but significant correlations with the theory component (r = 0.331 and 0.410, respectively).

The clinical component exhibited a high correlation with OSCE (0.888). Both clinical and OSCE components correlated well with the theory component (0.518 and 0.551). The Picture test also showed a strong correlation with the clinical component (0.810).

A weak and non-significant positive correlation was found between Progressive assessment and Picture test (0.158, p=0.066) while a weak but significant correlation was noted between SAQ and Orals (r=0.220, p=0.010).

Table 2: Correlation between components of final year medical students' psychiatry exam [N=137]

		Progressive Assessment **(PA)**	Final MBBS **MCQ**	Final MBBS **SAQ**	Final MBBS **ORAL**	Final MBBS **Theory**	Final MBBS **Picture Test**	Final MBBS **OSCE**	Final MBBS **Clinicals**
Progressive assessment **(PA)**	Pearson Correlation	1							
	p-value								
	N	137							
Final MBBS **MCQ**	Pearson Correlation	0.507**	1						
	p-value	**< 0.001**							
	N	137	137						
Final MBBS **SAQ**	Pearson Correlation	0.322**	0.351**	1					
	p-value	**< 0.001**	**< 0.001**						
	N	137	137	137					
Final MBBS **ORAL**	Pearson Correlation	-0.231**	0.309**	0.220**	1				
	p-value	**0.007**	**< 0.001**	**0.010**					
	N	137	137	137	137				
Final MBBS **Theory**	Pearson Correlation	0.531**	0.944**	0.641**	0.331**	1			
	p-value	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**				
	N	137	137	137	137	137			
Final MBBS **Picture Test**	Pearson Correlation	0.158	0.374**	0.200*	0.300**	0.410**	1		
	p-value	0.066	**< 0.001**	**0.019**	**< 0.001**	**< 0.001**			
	N	137	137	137	137	137	137		
Final MBBS **OSCE**	Pearson Correlation	0.357**	0.445**	0.339**	0.354**	0.518**	0.450**	1	
	p-value	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		
	N	137	137	137	137	137	137	137	
Final MBBS **Clinical**	Pearson Correlation	0.315**	0.485**	0.326**	0.386**	0.551**	0.810**	0.888**	1
	p-value	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	
	N	137	137	137	137	137	137	137	137

** *Significant at p<0.001, * Significant at p<0.05*

Table 3 presented a comparison of theory and clinical examination results within male and female groups using a paired sample t-test. The analysis indicated no significant within-gender differences between PA and final MBBS total scores (p>0.05 for both genders). However, females had higher overall scores in both PA (64.69 vs 61.06) and final MBBS total (64.09 vs 61.31).

Among male students, the mean score for PA (formative assessment) was 61.06, nearly identical to the final MBBS total score (summative), which was 61.31, with no significant difference observed (t = 0.345, p = 0.731). Similarly, female students exhibited comparable performance in PA, with a mean score of

64.69, and final MBBS total score, with a mean score of 64.09, with no significant difference observed (t = 0.739, p = 0.461).

While the analysis did not reveal significant within-gender differences between formative and summative assessments, it is noteworthy that female students scored higher on average than male students in both assessments.

Further analysis to determine whether these between-gender differences are statistically significant using Cohen's d to calculate effect size showed a large effect size for progressive assessment (0.85) and a moderate effect size (0.61) for the final MB score.

Table 3: Comparison of Theory and Clinical examination results within male and female gender using paired sample t test.

Variables	Gender	Mean	Std	t-test	p-value
Male	PA	61.06	4.03	0.345	0.731
	Final MBBS total	61.31	4.92		
Female	PA	64.69	4.46	0.739	0.461
	Final MBBS total	64.09	4.43		

Table 4 presented statistically significant gender differences in final MBBS performance (p=0.001 for both theory and clinical components). Female students had higher scores in both domains, achieving higher means in theory (32.24±2.82 compared to males' 30.53±2.69) and clinical assessments (31.86±2.37 compared to males' 30.78±2.85). The smaller standard deviations among female students, particularly in clinical examination (2.37 vs. 2.85), indicated more consistent performance patterns compared to male students.

Further analysis using Cohen's d to calculate effect size for theory and clinical examination indicated a moderate effect (0.63) for theory and a moderate but lower effect (0.42) for clinicals.

Table 4: Comparison of theory and clinical examination results between gender in final MBBS examination

Variables	Gender	Mean	Std	t-test	p-value
Theory	Male	30.53	2.69	3.616	**0.001***
	Female	32.24	2.82		
Clinicals	Male	30.78	2.85	-2.367	**0.019***
	Female	31.86	2.37		
**Significant at p<0.05*					

Discussion

The findings of this study provide significant insights into the reliability of both formative and summative assessments in medical education, variations in performance according to gender, and the interplay between theoretical and clinical examinations. These results contribute to the overall discourse on assessment validity, fairness in medical education, and the optimization of evaluation instruments.

The high internal reliability of the combined assessments in our study (α = 0.857) demonstrates the robustness of the assessment system, aligning with best practices in medical education that advocate for multifaceted assessment approaches [24,25].

However, the lower reliability among female students (α = 0.553) when formative and summative assessments were combined, compared to males (α = 0.618), suggests potential gender-related biases or differences in response patterns. This disparity may reflect variability in interaction with formative feedback or alignment of our assessment methods with gendered learning preferences [2].

The moderate reliability of the final examination alone (α = 0.778) confirms previous studies indicating that while written examinations are sufficient, their validity is enhanced by supportive workplace-based assessments [27].

The positive correlations between progressive assessments (PA) and theory components (MCQ: r = 0.507; SAQ: r = 0.322) suggest that formative assessments are good predictors of summative performance, supporting their inclusion within competency-based models [27].

The weak correlation between PA and the picture test (r = 0.158, p = 0.066) raises questions about the utility of the assessment tools, resonating with criticism that these approaches may not align well with the broader competency measurements [28].

Progressive Assessment (PA) correlates moderately with final MBBS written examination but poorly with Final MBBS picture test, suggesting that PA tests knowledge more than clinical applications when compared to picture test.

The extremely high correlation between clinical performance and OSCE (r = 0.888) underscores the validity of simulation-based assessments. This is comparable to what is seen in surgical training where OSCEs show how well one can do one's job [27], however, no within gender differences between Pas and Final MBBS suggests PA reliably predicts final performance.

Concerning gender, the present study show that female students outperformed male students in both theory and clinical examinations, with moderate effect sizes (Cohen's d = 0.63 and 0.42, respectively). This finding aligns with global trends in STEM education, where females perform as well as or better than males when provided with equal opportunities [29,30].

However, the lower reliability found in combined tests for females in this study, suggests that current assessment methods may not fully capture their competencies, due to biases in formative feedback systems or test designs. The lower standard deviations for females (2.37 vs. 2.85 for clinicals) also show that their performance patterns are more constant. This could mean that they are better at self-regulating or sticking to regular learning routines [29].

Studies show females score higher in communication stations (e.g, empathy, patient counseling) while males may excel in procedural stations (e.g., suturing, CPR), which emphasize technical and manual skills, where they may have greater confidence or aptitude [31].This may explain why females performed better, as psychiatry examinations require more enhanced communication skills such as interpersonal interactions and empathy [32-34].Other possible explanation is that females are more likely to engage in collaborative learning and seek feedback, enhancing clinical skill development than their male counterparts [19,35,36].

Furthermore, research shows that female may outperform male students in lower-stakes assessments while this trend can reverse in high-stakes assessments. For instance, females tend to perform better in progressive assessments but may struggle under higher pressure during final summative examinations to

evaluate anxiety experienced in such high-pressure situations [37]. This may explain why males did better in the final MBBS compared to their performance in progressive assessment.

The strength of this study lies in the high internal consistency of all the components of the assessment suggesting well-constructed questions and good psychometric properties. The comprehensive assessment mix/multimodal approach reduces bias to an extent. The strong correlation between theory and clinical performance suggests integrated learning. Valuable gender comparison also helps improve the strength of this study. Females outperforming males aligns with previous literatures [19,21,22,37]. However, no within-gender difference between the PA and Final MB suggests PA reliably predicts final performance.

The findings in this study have wide implications for Medical Education, especially in our clime where resources are limited. One is the need for more refining of formative assessments. The gender gap in reliability calls for a critical review of formative tools to ensure that they are equitable and aligned with summative goals. As recommended in competency-based frameworks, incorporating iterative feedback cycles could lessen these disparities [27].

There is also a need for evaluations that take gender into account. Drawing on UNESCO's advocacy for inclusive educational tools, institutions should explore whether specific assessment methods (e.g., picture tests) unintentionally disadvantage certain groups [29].

It also calls for the triangulation of data. Studies that criticize relying too much on traditional assessments suggest that evaluations should integrate multiple measures to enhance validity alongside these traditional methods (e.g., peer reviews, self-assessments) [38].

Limitations and Future Directions

The single-institution design limits the generalizability of the findings from this study, a challenge noted in broader assessments of medical education frameworks [1]. Future research should explore longitudinal trends across diverse cohorts and institutions. The addition of qualitative surveys to explore how cultural and curricular factors may influence gendered experiences in medical assessments, may further clarify the underlying reasons for discrepancies in reliability, thereby enhancing the insights gained from this quantitative research [39].

Conclusion

This study highlights the nuanced interplay between assessment reliability, gender, and performance in medical education. While the evaluation system demonstrates strong validity overall, the gendered differences in reliability and performance highlight the need for more equitable assessment practices. By addressing these gaps, medical educators can foster a more inclusive and robust training environment, enhancing both learner outcomes and patient care.

References

1. Harden RM. Outcome-based education: The future is today. Med Teach. 2007;29(7):625-9.

2. Kordestani Moghaddam A, Khankeh HR, Shariati M, Norcini J, Jalili M. Educational impact of assessment on medical students' learning at Tehran University of Medical Sciences: a qualitative study. BMJ Open. 2019;9(7): e031014.

3. Harden RM, Crosby JR. AMEE Guide No 20: The good teacher is more than a lecturer—the twelve roles of the teacher. Med Teach. 2000;22(4):334-47.

4. Fallatah HI, Tekian A, Park YS, Al Shawa L. The validity and reliability of the sixth-year internal medical examination administered at the King Abdulaziz University Medical College. BMC Med Educ. 2015; 15:10.

5. Preston R, Gratani M, Owens K, Roche P, Zimanyi M, Malau-Aduli B. Exploring the impact of assessment on medical students' learning. Assess Eval High Educ. 2019;45(1):109-24.

6. Safari M, Shahidi S, Avizhgan M. Challenges in applying theoretical knowledge from the general medicine course in clinical practice: a qualitative study. J Med Educ Dev. 2024;17(54):1-10.

7. Schildkamp K, van der Kleij FM, Heitink MC, Kippers WB, Veldkamp BP. Formative assessment: a systematic review of critical teacher prerequisites for classroom practice. Int J Educ Res. 2020; 103:101602.

8. Ismail SM, Rahul DR, Patra I, Rezvani E. Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. Lang Test Asia. 2022;12(1).

9. Hawe E, Dixon H. Assessment for learning: a catalyst for student self-regulation. Assess Eval High Educ. 2017;42(8):1257-67.

10. Dolin J, Black P, Harlen W, Tiberghien A. Exploring Relations Between Formative and Summative Assessment. In: Dolin J, Evans R, editors. Transforming Assessment. Cham: Springer International Publishing; 2018. p. 53-80.
11. Bloom BS, Engelhart MD, FurstEJ, Hill WH, Krathwohl DR. Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive domain. New York: David McKay; 1956.

12. Anderson LW, Krathwohl DR. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Longman; 2001.

13. Tuma F, Nassar AK. Applying Bloom's taxonomy in clinical surgery: practical examples. Ann Med Surg (Lond). 2021; 69:102656.

14. Riskiyana R, Qomariyah N, Hidayah RN, Claramita M. Towards improving soft skills of medical education in the 21st century: a literature review. Int J Eval Res Educ. 2022;11(4):2174-84.

15. MunteanLM, Nireștean A, Sima-Comaniciu A, Mărușteri M, Zăgan CA, Lukacs E. The relationship between personality, motivation and academic performance at medical students from Romania. Int J Environ Res Public Health. 2022;19(15):8993.

16. Al Mamun MMA, Alam KK, Talukder HK, Jahan S, Islam N, Chowdhury M. Medical students' motivational factors and academic performance – undergraduate level of Bangladesh. Bangladesh J Med Educ. 2024;15(1):67-74.

17. Hayat AA, Salehi A, Kojuri J. Medical student's academic performance: The role of academic emotions and motivation. J Adv Med Educ Prof. 2018;6(4):168-75.

18. Henning MA, Krägeloh CU, Moir F, Webster CS. The relationship between coping strategies and student performance in medical education: a longitudinal study. Med Educ Online. 2015; 20:27576.

19. Saxena S, Wright WS, Khalil MK. Gender differences in learning and study strategies impact medical students' preclinical and USMLE step 1 examination performance. BMC Med Educ. 2024; 24:504.

20. Li B, Jacob-Brassard J, Dossa F, Salata K, Kishibe T, Greco E, et al. Gender differences in faculty rank among academic physicians: a systematic review and meta-analysis. BMJ Open. 2021;11(11): e050322.

21. Kibuuka R, Mpasa F, Atuhairwe I, Agaba B, Nakattudde P, Owusu-Sekyere S, et al. Factors associated with retakes in health Professions Courses: A case study of Five selected Universities in Sub-Saharan Africa. Res Sq [Preprint]. 2024: rs.3.rs-5368416.

22. Blanch DC, Hall JA, Roter DL, Frankel RM. Medical student gender and issues of confidence. Patient EducCouns. 2008;72(3):374-81.

23. Nguyen M, Zhao W, Olsen C. Gender differences in study strategies and performance among medical students: a LASSI-based study. BMC Med Educ. 2024; 24:149.

24. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Med Teach. 2010;32(8):676-82.

25. Song X, Cleaves E, Gluzman E, Kotlyar B, Russo RA, Schilling DC, et al. A Scoping Review of Assessments in Undergraduate Medical Education: Implications for Residency Programs and Medical Schools. Acad Psychiatry. 2025;49(3):263-73.

26. Preston R, Gratani M, Owens K, Roche P, Zimanyi M, Malau-Aduli B. Exploring the Impact of Assessment on Medical Students' Learning. Assess Eval High Educ. 2019;45(1):109-24.

27. The Council on Undergraduate Research. Assessment Tools [Internet]. 2024 [cited 2025 Jun 29]. Availablefrom: https://www.cur.org/resources-publications/mentor-resources/assessment-tools/

28. Constantinou C, Wijnen-Meijer M. Student evaluations of teaching and the development of a comprehensive measure of teaching effectiveness for medical schools. BMC Med Educ. 2022;22(1):113.

29. LibGuides: Undergraduate Research: Studies on the Impact of Undergraduate Research on Students [Internet]. [cited 2025 Jun 29]. Available from: https://libguides.usu.edu/c.php?g=1066148&p=7820635

30. Kaifi R, Alshamrani K, Al-Nasser S, Omair A, Althaqafy M. The Relation Between Theoretical and Practical Exams for Health Sciences Students at King Saud Bin Abdulaziz University for Health Sciences- Jeddah. Adv Med Educ Pract. 2024; 15:419-30.

31. Bhatti U, Chohan MN, Effendi S, Khidri FF, Ahmadani R, Ghaffar B. Gender-Based Variation in The Academic Performance of Medical Students: A Cross-Sectional Study. BioSight. 2023;4(2-3):45-50.

32. Novais F, Ganança L, Barbosa M, Telles-Correia D. Communication skills in psychiatry for undergraduate students: a scoping review. Front Psychiatry. 2022; 13:972703.

33. Ross J, Watling C. Use of empathy in psychiatric practice: constructivist grounded theory study. BJ Psych Open. 2017;3(1):26-33.

34. Fuller M, de Jong M, Kamans E, Wolfensberger M, van Vuuren M. Empathy competencies and behaviours in professional communication interactions: self-versus client assessments. Bus Prof Commun Q. 2023;86(2):167-86.

35. Mohammed G, Bennett D. Effect of gender on medical students' clinical learning experience in a multicultural setting. Adv Biomed Health Sci. 2023;2(3):131-8.

36. Samuriwo R, Patel Y, Webb K, Bullock A. 'Man up': medical students' perceptions of gender and learning in clinical practice: a qualitative study. Med Educ. 2020;54(2):150-61.

37. Montolio D, Taberner PA. Gender differences under test pressure and their impact on academic performance: a quasi-experimental design. J Econ Behav Organ. 2021; 191:1065-90.

38. Stanford University. Course feedback as a measure of teaching effectiveness [Internet]. 2015 [cited 2025 Jun 29]. Available from: https://evals.stanford.edu/evaluating-teaching/course-feedback-measure-teaching-effectiveness

39. Søndergaard MK, Degn L, Lassesen B. Thinking or talking about teaching? Student evaluation as an occasion for dialogue or reflection on teaching. Tert Educ Manag. 2024;30(2):147-63.