Medical Students' Performances Using Different Assessment Methods during the Final Examination in Internal Medicine at the University of Benghazi, Libya

Najat Buzaid[1,2]   Sami A. Lawgaly[1,3]   Saleh M. Alawgali[1,2]   Amina Albash[1,2]   Mousa Alfakhri[1,2]

[1] Department of Internal Medicine, Faculty of Medicine, University of Benghazi, Libya

[2] Department of Medicine, 7th of October Hospital, Libya

[3] Department of Medicine, Benghazi Medical Center, Benghazi, Libya

Address for correspondence Sami A. Lawgaly, AlSilini Street, Boatney, Benghazi, Libya (e-mail: Sami.lawgaly@uob.edu.ly).

Abstract

Background   Distinctive evaluation tools assess diverse fields of learning that considerably impact the learning process.

Objective   To compare and correlate the performances of undergraduate final year medical students in written, clinical, and viva examinations in the subject of internal medicine.

Methods   This is a retrospective study. After authority approval, data was collected from final year examination results during 2019 to 2020. All the students of the medical school at University of Benghazi were included in this study. Their gender and their written, clinical, viva, and total scores were included. Data were coded and transferred from Excel to SPSS version 24 and expressed as frequencies and percentages. Chi-squared analysis was performed to test for differences in the proportions of categorical variables between two or more groups. Odd ratio (OR) is used to calculate the odds of passing the subject based on scores in different types of exams. Person's correlation (R) is used to evaluate the consistency of students' performances in different examinations. A p-value of less than 0.05 was considered the cut-off value of significant.

Results   The total number of students was 679, out of which 499 (73.5%) were females and 180 (26.5%) were males. The total number of students who passed the course was 422 (62%) with no significant differences between males and females. A statistically significant ($p < 0.001$) greater percentage of students achieved a passing score in clinical assessment (502 [73.9%]), followed by viva assessment (458.0 [67.5%]). The students performed the worse in written examination with only 291/679 (43%) students passing the examination, with no gender-based differences. There was a highly significant association between the total score of students who passed the subject and their scores in the written examination with an OR of 2.3 ($p < 0.001$). Viva examination and total score OR was 0.79 with no significant differences for males or females. On the contrary, there was a statistically significant negative association

Keywords

► medical
► performance
► clinical examination
► written examination
► viva examination
► internal medicine
► Benghazi
► Libya

between clinical exams and total scores of students who passed the subject (OR = 0.58). There was a highly significant correlation ($p < 0.001$) between written examination and viva examination ($R = 0.638$), between written examination and clinical examination ($R = 0.629$), and between clinical and viva examinations ($R = 0.763$).

Conclusion Students demonstrated higher performance on clinical and viva exams compared with written exams. Additionally, there were no notable disparities in results between male and female students across any of the three exam types. The written exam served as the most reliable indicator of a student's success in the subject. Furthermore, the data revealed a positive correlation between scores on the different exam formats, indicating that students exhibited consistent performance across all modes of evaluation.

ملخص المقال باللغة العربية

أداء طلاب الطب أثناء الامتحان النهائي في الطب الباطني باستخدام أساليب التقييم المختلفة بجامعة بنغازي، ليبيا.

المؤلفون: نجاة بوزيد، سامي الأوجلي، صالح الأوجلي، أمينة الباش، موسى الفاخري، قسم الباطنة جامعة بنغازي، ليبيا.

المؤلف المسؤول: سامي الأوجلي البريد الإلكتروني sami.lawgaly@uob.edu.ly -

الخلفية: تختلف أدوات التقييم المستعملة في تقييم مجالات التعلم المتنوعة التي تؤثر بشكل كبير على عملية التعلم.

الهدف: مقارنة ودراسة العلاقة في أداء طلاب السنة النهائية في الطب في الامتحانات التحريرية والسريرية والشفوية في مادة الطب الباطني.

الطرق: هذه دراسة بأثر رجعي. تم جمع البيانات من نتائج امتحانات السنة النهائية خلال العام 2019-2020م بعد موافقة الكلية. تم تضمين جميع الطلاب في هذه الدراسة، شاملة جنس الطالب، الدرجات التي تحصل عليها في الامتحانات التحريرية والسريرية والشفوية، وكذلك مجموع الدرجات. تم ترميز البيانات ونقلها من Excel إلى SPSS الإصدار 24 وتم التعبير عنها بالتكرارات والنسب المئوية. تم إجراء تحليل مربع كاي لاختبار الاختلافات في نسب المتغيرات الفئوية بين مجموعتين أو أكثر. كما تم استخدام نسبة الأرجحية (OR) لحساب احتمالات النجاح في المادة بناءً على الدرجات التي تحصل عليها الطالب في الأنواع المختلفة من الاختبارات. وتم استخدام معامل الارتباط (R) لتقييم مدى ارتباط أداء الطلاب في الامتحانات المختلفة. تم اعتبار مستوى 0.05 >P القيمة الفاصلة ذات الأهمية.

النتائج: بلغ إجمالي عدد الطلاب 679 طالباً، منهم 499 (73.5%) إناث و180 (26.5%) ذكور. وبلغ إجمالي عدد الطلاب الذين اجتازوا المقرر 422 طالباً بنسبة 62%، ولا توجد فروق ذات دلالة إحصائية بين الذكور والإناث. حققت نسبة كبيرة من الطلاب درجة النجاح في التقييم السرير (73.9%)، يليها التقييم الشفوي (67.5%)، في حين كان الامتحان الكتابي هو أدنى أداء (43%)، مع عدم وجود فارق بين الجنسين. وكانت هناك علاقة ذات دلالة إحصائية عالية بين مجموع درجات الطلاب الذين اجتازوا المادة ودرجاتهم في الامتحان التحريري حيث كانت نسبة الأرجحية تساوي 2.3. في حين أن نسبة الأرجحية بين مجموع الدرجات والامتحان الشفوي كانت 0.79 مع عدم وجود فروق ذات دلالة إحصائية بين الجنسين. وعلى العكس من ذلك، كانت هناك علاقة سلبية ذات دلالة إحصائية بين كل من الامتحانات السريرية ومجموع درجات الطلاب الناجحين في المادة (نسبة الأرجحية 0.58). كما كان هناك معامل ارتباط عال ذو دلالة إحصائية ما بين كل من الاختبار التحريري والشفوي (0.638) وما بين الاختبار التحريري والسريري (0.629)، وكذلك بين اختبار السريري والشفوي (0.763).

الاستنتاج: أظهر الطلاب أداءً أعلى في الاختبارات السريرية والشفوية مقارنة بالاختبارات التحريرية. بالإضافة إلى ذلك، لم تكن هناك اختلافات ملحوظة في النتائج بين الطلاب والطالبات في أي من أنواع الاختبارات الثلاثة. كان الاختبار التحريري المؤشر الأكثر موثوقية لنجاح الطالب في هذا المقرر. علاوة على ذلك، كشفت النتائج عن وجود معامل ارتباط عال بين الدرجات في الاختبارات المختلفة، مما يشير إلى أن الطلاب أظهروا أداءً ثابتًا في جميع طرق التقييم.

الكلمات المفتاحية: طبي، الأداء، الاختبار السريري، الاختبار التحريري، الاختبار الشفوي، الطب الباطني، بنغازي، ليبيا.

Introduction

Assessment is the most important factor that drives students' learning, as students tend to study materials that will be assessed. Bloom's taxonomy was originally proposed by Benjamin Bloom in 1956 and has since been revised. The taxonomy consists of six levels: remembering, understanding, applying, analyzing, evaluating, and creating. Each level builds upon the previous one and requires a higher level of cognitive skill.[1] There are different methods of assessment that examine different domains of Bloom's taxonomy. Theory essays test the knowledge (level 1), at this stage, there will be an assess- ment of how will the student learn new knowledge. Questions that contain verbs like explain and compare will test comprehension (level 2). Exams that instruct students to apply and compare represent level 3, while those that test analysis and synthesis represent levels 4 and 5. Finally, level 6 tests evaluation and conclusion. Written examinations usually test levels 1 to 3, while clinical examinations test levels 2 to 6.[2] On the other hand, to create a competent graduate, other skills should be evaluated; like communication, analytical skills, teamwork skills, and evidence-based medical care.[3,4,6] Students' assessments can be performed by many methods including short essay questions, students' projects, short and long case

assessments, objective structured clinical examination (OSCE), matching, multiple choice questions (MCQ), portfolios, use of patient simulators, video assessment, self-assessment, reports, audits, oral exams, logbook, and peer assessment.[7–9] The choice of assessment methods depends on the domains being tested. Different learning outcomes should be tested by suitable assessment tools. Usually, a combination of assessment methods is required to test different learning outcomes, and good assessment methods will ultimately promote students' learning. Clinical competency is usually assessed by OSCE, mini-clinical evaluation exercises, directly observed procedural skills, and short answer questions.[9–11] In this retrospective analysis, we shall examine the academic achievements of students from the medical school at Benghazi University in the field of internal medicine, by employing three distinct forms of examination.

Methods

This is a retrospective study. After authority approval, data was collected from final year examination results during the year 2019–2020. All students were included in the study. Students' gender, written, clinical, viva, and total scores in the subject were included. The written examination is composed of two papers; each paper with 50 questions; paper 1 included 50 case scenarios with multiple choice questions and paper 2 included 50 multiple choice questions. Clinical examination is composed of five stations: four clinical and one viva examination station.

The total score for the final year examination was 300. Scores were distributed as follows; 100 marks for the written examination, 150 marks for the clinical exam, and 50 marks for the viva examination. The required pass score percentage was 60%; which means 180 marks for the total score, 60 marks for the written exam, 90 marks for the clinical examination, and 30 marks for *the viva* exam.

Statistical Methods

Data was coded and transferred from Excel to Statistical Package for Social Sciences (SPSS) version 24 (Chicago, IL, United States). The data included the number of students who passed or failed each type of the three exams according to gender. Data was expressed as frequency (percentage). Chi-square analysis was performed to test for the differences between two or more groups. Odd ratio (OR) is used to calculate the odds of passing the subject based on scores in different types of exams. Person's

Table 1 Frequency and passing rates by gender in internal medicine

Gender	Frequency (%)		Passed		Failed	
	Number	%	Number	%[a]	Number	%
Female	499	73.5	314	62.9	185	37.1
Male	180	26.5	108	60	72	40
Total	679	100	422	62.2	257	37.9

[a]Percentage calculated from the total number of students in the respective gender.

correlation (R) is used to evaluate the consistency of students' performances in different exams. The level of $P < 0.05$ was considered the cut-off value of significance.

Results

General Characteristics
The total number of students was 679, 499 (73.5%) were females and 180 (26.5%) were males. The total number of students who passed the subject was 422 (62.2% of the total number of students), 314 were females (62.9% of the total number of females) and 108 were males (60% of the total number of males) with no significant difference between male and female students (►**Table 1**).

Comparison of Students' Performance in the Written, Clinical, and Viva Examinations
In total, students had a higher performance in clinical examination (73.9%) and viva examination (67.5%) compared with written examination (43%). This was statistically significant ($p < 0.001$) and applied to both male and female students. However, there were no significant differences in performance between male and female students in any of the three types of exams.

The Odd Ratio of the Relationship between the Total Score and Scores in Different Exams
►**Table 2** shows the OR between the total score of students who passed the subject and their scores in the written examination, indicating a statistically significant positive association ($p < 0.001$). This means that the odds of passing the final exam were about two times higher than passing a written exam with a high degree of significance for both males and females. On the other hand, the viva examination OR with the total score was 0.79 with no significant differences for males or

Table 2 Odds ratio of total score compared with scores of different types of examinations with gender-based analysis

	Gender	Passed written examination			Passed viva examination			Passed clinical examination		
		OR	95% CI	*p*	(OR)	95% CI	*p*	OR	95% CI	*p*
Passed the subject	Female	2.3	1.8–3	0.0001	0.77	0.59–1.00	N.S.	0.57	0.43–0.75	0.001
	Male	1.9	1.2–2.8	0.003	0.85	0.55–1.30	N.S.	0.59	0.39–0.92	0.02
	Total	2.2	1.7–2.7	0.0001	0.79	0.63–0.99	0.04	0.58	0.46–0.73	0.001

Abbreviations: CI, confidence intervals; OR, odds ratio.

Table 3 Correlation between different types of examinations used to evaluate students' performance in internal medicine

		Written examination marks	Clinical examination marks	Viva voce marks
Written examination marks	Pearson's correlation	1	0.629	0.638
	Sig. (2-tailed)		0.000	0.000
	N	679	679	679
Clinical examination marks	Pearson's correlation	0.629	1	0.763
	Sig. (2-tailed)	0.000		0.000
	N	679	679	679
Viva voce marks	Pearson's correlation	0.638	0.763	1
	Sig. (2-tailed)	0.000	0.000	
	N	679	679	679

females. On the contrary, we found a statistically significant negative association between total scores and clinical exams of students who passed the subject (OR = 0.58).

Correlations between Students' Performances in the Different Types of Exams

The relationship between the different types of exams used to evaluate students' performance was calculated by using Pearson's correlation (►**Table 3**). It shows that there is a highly significant correlation between the different types of examinations, with the lowest between clinical and written exams ($R = 0.629$) and the highest between viva and clinical exams ($R = 0.763$).

Discussion

The type of assessment method can influence students' learning.[11–13] Different assessment tools are evaluating different domains of learning.[14] In this study, we are reporting the results of students in internal medicine using three different types of exams, each evaluating different domains of learning (refer to the Introduction). The total number of students who passed the subject was 62.2% with nearly equal percentages for females (62.9%) and males (60%; ►**Table 1**), indicating that gender has no effect on the total performances of students in this study. Similar results were reported where no significant differences were found in the performance of males and females on preclerkship OSCEs or Essentials of Clinical Medicine semester final

exams.[15] However, our results are different from other studies, where they found gender differences regarding students' performances; one study showed a better male performance,[16] while another two studies showed a better female performance.[17,18]

When we looked at the performances of students in different types of exams (►**Table 4**), we found that both male and female students performed better in clinical and viva examinations than in written examinations, with no significant differences in performance between male and female students in any of the three types of exams. Similar results were reported in another study, where scores of the clinical examination were significantly higher than the written examination.[19] There could be several reasons why students performed differently in the three types of exams. Clinical and viva examinations are typically more interactive and require students to apply their knowledge in practical situations, which may better reflect their understanding of the material. On the other hand, written exams may be more focused on testing memorization and recall of information, which may not necessarily reflect a student's ability to apply that knowledge. Additionally, the format of the exams and the types of questions asked may also contribute to differences in performance. Another explanation is that the clinical assessment tools used in our study might be not completely objective and examiner factors could play a role. The suggested sensible solution for this inconsistency is the use of Objective Structured Clinical Examination (OSCE) which was proven in many studies to be a reliable and valid mode of assessment for clinical skills as

Table 4 Comparison of student performance in written, clinical, and viva examinations with gender-based analysis

Gender	Frequency	Written examination		Viva examination		Clinical examination		Chi-square statistic (p)
		Number	%[a]	Number	%[a]	Number	%[a]	
Female	499.0	211.0	42.3	343.0	68.7	373.0	74.7	0.001
Male	180.0	80.0	44.4	115.0	63.9	129.0	71.7	0.001
Total	679.0	291.0	42.9	458.0	67.5	502.0	73.9	0.001

[a]Percentage calculated from the total number of students in the respective gender.

students obtained less scores on the OSCE than the traditional clinical examination.[19–21]

The OR statistics were used to calculate the odds of passing the subject based on scores in different types of exams, and which type of examination can be the best predictor of students passing in the subject of internal medicine. ►Table 2 shows that there was a statistically significant positive association between the total marks of students and their marks in the written exams. Specifically, the OR was about two times higher for passing the final exam compared with passing a written exam, and this difference was highly significant ($p < 0.001$). This suggests that students who perform well on the written exam are more likely to pass the subject, and this relationship holds true for both male and female students. In other words, doing well on the written exam is a good predictor of success in our subjects, and this finding is statistically significant. On the other hand, the OR between the viva examination score and the total score was 0.79, which suggests a weaker relationship compared with the written exam and the total score. This means that performing well on the viva examination is not as good a predictor of success as the written exam. On the opposite side, the study found a statistically significant negative association between the total score and clinical exam scores of students who passed the subject. Specifically, OR was 0.58. This finding is in contrast to the positive association found between the written exam and the total score. It suggests that clinical exams may be a weaker predictor of success on the total score compared with the written exam. Interestingly, our study found no significant differences in performance between male and female students in any of the three types of examinations. This means that gender did not have a significant effect on student performance, contrary to the "gender gap" reported in some literature.[15]

To evaluate the performance of students, we used Pearson's correlation to calculate the relationship between their scores in various types of exams (as shown in ►Table 3). The results indicate a strong correlation between different exams, indicating consistency in student performance across exams. This means that if a student performs well in one exam, they are likely to perform well in the others too. On the other hand, if a student performs poorly in one exam, they are also likely to perform poorly in the other exams.

Conclusion

This study found that students performed better in the clinical and viva examinations than in the written examination. There was no gender difference in the performance of male and female students across the three types of exams. The written exam was the strongest predictor of student success in the subject. The student's performance was consistent in the three types of exams and not affected by gender.

Limitations of This Study

The study only included students from one subject and one batch, which may limit the generalizability of the findings

to other contexts and domains. Moreover, the study did not control for other factors that may affect student performance, such as motivation, prior knowledge, learning styles, or instructor quality. Additionally, the study did not use specific outcomes or competencies to evaluate student performance, but rather a general score that may not capture the nuances of student learning and achievement.

Conflict of Interest
None declared.

References

1 Adams NE. Bloom's taxonomy of cognitive learning objectives. J Med Libr Assoc 2015;103(03):152–153
2 Anderson LW, Krathwohl DR. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York, NY: Longman; 2021
3 Case SM, Swanson DB. Constructing Written Test Questions for the Basic and Clinical Sciences. 3rd. ed. Philadelphia, PA: National Board of Medical Examiners; 2002
4 Day SC, Norcini JJ, Diserens D, et al. The validity of the essay test of clinical judgement. Acad Med 1990;65(09):S39–S40
5 Epstein RM, Hundert EM. Defining and assessing clinical competence. JAMA 2002;387:226–235
6 Boulet JR, Rebbecchi TA, Denton EC, McKinley DW, Whelan GP. Assessing the written communication skills of medical school graduates. Adv Health Sci Educ Theory Pract 2004;9(01):47–60
7 Norcini JJ. Peer assessment of competence. Med Educ 2003;37 (06):539–543
8 Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990;65(9, Suppl):S63–S67
9 Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. Ann Intern Med 2003;138(06):476–481
10 Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001;357(9260):945–949
11 Epstein RM. Assessment in medical education. N Engl J Med 2007; 356(04):387–396
12 Kerdijk W, Snoek JW, van Hell EA, Cohen-Schotanus J. The effect of implementing undergraduate competency-based medical education on students' knowledge acquisition, clinical performance and perceived preparedness for practice: a comparative study. BMC Med Educ 2013;13:76
13 Amin TT, Kaliyadan F, Al-Muhaidib NS. Medical students' assessment preferences at King Faisal University, Saudi Arabia. Adv Med Educ Pract 2011;2:95–103
14 Charles J, Kalpana S, Stephen Max LJ, Shantharam D. A cross sectional study on domain based evaluation of medical students. IOSR J Res Method Educ 2014;4(04):33–36
15 Hesse DW, Ramsey LM, Bruner LP, et al. Exploring academic performance of medical students in an integrated hybrid curriculum by gender. Med Sci Educ 2023;33(02):353–357
16 Conger D, Long MC. Why are men falling behind? Gender gaps in college performance and persistence. Accessed April 2, 2015 at: http://www.texastop10.princeton.edu
17 Maliki AE, Ngban AN, Ibu JE. Analysis of students' performance in Junior Secondary School Mathematics Examination in Bayelsa State of Nigeria. Stud Home Comm Sci 2009;3:131–134
18 Alam KK, Begum SN, Nargis T. Feedback on formative assessment in undergraduate medical education Bangladesh. Bang J Physiol Pharmacol 2009;25(1–2):18–22

19 Memon S, Shaikh SU. Comparison of performance on written and OSCE assessment during end semester pediatric examination. Pak J Med Sci 2020;36(04):711–716

20 Rahman N, Ferdousi S, Hoq N, Amin R, Kabir J. Evaluation of objective structured practical examination and traditional practical examination. Mymensingh Med J 2007;16(01):7–11

21 Mondal R, Sarkar S, Nandi M, Hazra A. Comparative analysis between objective structured clinical examination (OSCE) and conventional examination (CE) as a formative evaluation tool in Pediatrics in semester examination for final MBBS students. Kathmandu Univ Med J 2012;10(37):62–65