Comparison of Assessment Tools in Online and On-Campus Undergraduate Medical Examinations amidst COVID-19 Pandemic

Muhammad Ali Rabbani[1], Hamd Binte Shahab Syed[1] and Faiza Ikram[2]

[1]*Department of Anatomy, CMH Multan Institute of Medical Sciences (CIMS), Multan, Pakistan*
[2]*Department of Physiology, CMH Multan Institute of Medical Sciences (CIMS), Multan, Pakistan*

ABSTRACT

Objective: To compare the discriminating ability of online assessment tools in the module examinations during the COVID-19 pandemic.

Study Design: Descriptive study.

Place and Duration of Study: Department of Anatomy, CMH Multan Institute of Medical Sciences (CIMS), Multan, from 22 June to 09 July 2021.

Methodology: In the academic year 2020, the first module examinations of the 2nd-year MBBS anatomy course was conducted on-campus *via* multiple-choice questions, short answer questions, and *viva-voce*. Owing to the COVID-19 lockdown, the following two module examinations were held online. The effectiveness of different assessment tools in the three module examinations was compared by calculating the discrimination indices and the area under the curve (AUC), using the receiver operating characteristic (ROC) curve analysis.

Results: SAQs showed the highest discrimination index (0.38) compared to MCQs and *viva-voce* in the on-campus module examinations but dropped to the lowest in the subsequent online modules (0.24 and 0.18). In contrast, the discriminating power of the *viva-voce* increased from marginally (0.23) to very good (0.47 and 0.49) as the mode of assessment shifted to online examinations. The ROC AUC also showed the same pattern. In the second and third module examinations, the *viva-voce* had significantly higher ($p < 0.05$) AUC than MCQs alone and both the MCQs and SAQs, respectively.

Conclusion: *Viva-voce* had a significantly higher discriminating index than MCQs and SAQs in online examinations. On-campus, SAQs had the highest discriminatory index. Using this statistical approach, the effectiveness of different components of the online examinations can be monitored to improve the quality of online examinations amidst the COVID-19 crisis.

Key Words: *COVID-19, Distance learning, Online education, Discrimination index, ROC curve, Pakistan, Assessment.*

How to cite this article: Rabbani MA, Syed HBS, Ikram F. Comparison of Assessment Tools in Online and On-Campus Undergraduate Medical Examinations amidst COVID-19 Pandemic. *J Coll Physicians Surg Pak* 2022; **32(03)**:359-363.

INTRODUCTION

The coronavirus disease (COVID-19) was declared a pandemic by the World Health Organization in March 2020.[1] In addition to its innumerable healthcare, economic and social consequences, the COVID-19 pandemic also disrupted the education sector on a global scale.[2] Countless medical schools and universities halted their on-campus academic activities due to the government-imposed lockdown, leading to a massive paradigm shift in medical education towards online learning equivalents.[3]

Correspondence to: Dr. Muhammad Ali Rabbani, Department of Anatomy, CMH Multan Institute of Medical Sciences (CIMS), Multan, Pakistan
E-mail: ali-rabbani@hotmail.com

Received: July 10, 2021; Revised: September 02, 2021; Accepted: October 29, 2021
DOI: https://doi.org/10.29271/jcpsp.2022.03.359

During the early days of the lockdown, the primary focus was to assure the continuation of medical teaching remotely. This virtual learning was achieved by various methods, ranging from simply sharing written handouts and recorded lectures to the developing complex learning management systems, where synchronous teaching was carried out using live lectures and discussion forums.[4,5]

With the teaching methods in place and lockdown still in effect, the next big question was the conduction of online examinations. The computer-based examination is not a new concept in this time and age; and many institutes had already been implementing it all over the globe.[6] However, for the majority, the employment of such new strategies on such short notice posed a challenge for the faculty and students alike.

Institutes around the world came up with several creative solutions for solving this conundrum.[7,8] For example, Columbia University College of Physicians and Surgeons, New York, introduced the policy of un-proctored, open-book examinations to

their residents in surgical clerkship. The idea behind this was that the students would honour the code of conduct, and with a time limit coupled with specially made questions, the assessment will be reasonably neutral.[7] Other institutes introduced various strategies of remotely invigilated online examination (RIOE).[9] These varied from simple camera monitoring to using complex web-based proctoring software that monitored webcam, keystrokes, and on-screen activity, as the University of New England, Australia.[10] Moreover, in Pakistan, the government imposed a lockdown, rendering on-campus medical education impossible. Hence, the medical teaching and assessment were shifted to their online equivalents despite the lack of resources and previous exposure.

Compared to traditional campus-based examinations, the efficacy and reliability of these online assessments is still a topic of debate. Owing to the continually prevailing pandemic, the need of the hour is to devise some modalities to assess and improve the quality and integrity of online examinations. Researchers have devised various tools to assess this. Jaap and Elsalem utilised post-exam surveys from the students to analyse their experiences.[8,11] Other studies focused mainly on exam-results data analysis, comparing online to on-campus.[7,12] These, and most other such analyses, focused on one exam modality, *e.g.*, multiple-choice questions (MCQs); and then compared the mean scores achieved in various situations.[8,12]

Studies have compared the validity and discriminating ability of different exam components, *i.e.*, MCQs, SAQs, and *viva-voce* in on-campus exams.[13-15] However, to the best of the authors' knowledge, such studies have not been conducted to assess online examinations.

This study aimed to compare the discriminating ability of different exam components, *i.e.*, MCQs, SAQs, and *viva-voce*, in online and on-campus examinations. The authors intended to develop a statistical approach that can be utilised in the future online exams to monitor and improve the constructional validity of various examinations components best suited to available resources.

METHODOLOGY

This study was conducted at CMH Multan Institute of Medical Sciences, Multan, from June to July 2021. Second-year MBBS course was taught anatomy in three modules, followed by a summative exam at the end of each module. The modular exam consisted of theory multiple-choice questions (MCQs), short answer questions (SAQs) and *viva-voce* components. In the academic year 2020, the first modular exam was conducted traditionally on-campus. The theory exam was held in the examination hall under continuous invigilation. The *viva-voce* was conducted face-to-face by a single examiner over a period of three days. The second and third modules were conducted and assessed online, owing to the lockdown due to the COVID-19 pandemic. Theory exams were held online *via* Google forms with time-restricted access. The identity of the student attempting the exam was authenticated *via* video call on Zoom, followed by continuous video invigilation throughout the exam. However, no special software was used to invigilate or restrict their screen activity. Students gave *viva-voce* on a one-on-one video call on Zoom to the same examiner as their on-campus exam.

Data were collected retrospectively from exam scores of all three modular exams after approval from the Institutional Review Board and Ethics Committee (No. TW/51/CIMS). Percentage marks were obtained for MCQs, SAQs, and *viva* components. Total marks were calculated, giving the former three components equal weightage. Only those students of second-year MBBS, who had appeared in all three components of every module exam of the academic year 2020, were included in the study, *i.e.*, 73 out of 101 students.

The discrimination index (DI) of an item is defined as the degree to which it discriminates between the students with high and low scores.[16] It ranges between +1.00 and -1.00, with +1.00 being the case where all high scorers answer an item correctly, and none of the low scorers does. An exam item with a discrimination index of 0.40 and above is considered very good, 0.30-0.39 is reasonably good, 0.20-0.29 is marginal (subject to improvement), and 0.19 or less is a poor item.[3] For calculation of discrimination indices, in each module exam, the authors divided the students into high scorers (upper 27th percentile) and low scorers (lower 27th percentile), using Truman Kelley's "27% of the sample" group size.[17] Discrimination index of each item *i* (MCQs, SAQs & *viva*) was then calculated using the following formula:[18]

$$\textit{Discrimination index} = (\Sigma Hi - \Sigma L_i)/(N \times mi).$$

Where ΣHi is the sum of marks in the high scoring group in item *i*, ΣL_i is the sum of marks in the low scoring group in item *i*, *N* is the number of students in each group, and *mi* is the total marks in item *i*.

To validate further, the area under the receiver operating characteristic (ROC) curve of MCQs, SAQs, and *viva voce* was calculated by ROC analysis to signify their ability to predict and discriminate high and low achievers in the exam (here defined by the median total marks in that exam). The area under the curve (AUC) and 95% confidence interval was calculated from the ROC curve using achievement level as a binary outcome (high/low achievers). AUC ≥0.9 is considered outstanding discrimination, 0.8-0.9 is considered excellent, 0.7-0.8 is acceptable, 0.5-0.7 poor, and 0.5 suggests no discrimination.[19] Area under the curve (AUC) between the ROC curves of each exam modality were compared using ROC analysis in statistical package for social sciences (SPSS version 26).

RESULTS

In the first module exam conducted on-campus, SAQs had the highest and reasonably good (0.38) discrimination index compared with MCQs and *viva*, which had marginally good discrimination indexes of 0.27 and 0.23, respectively.

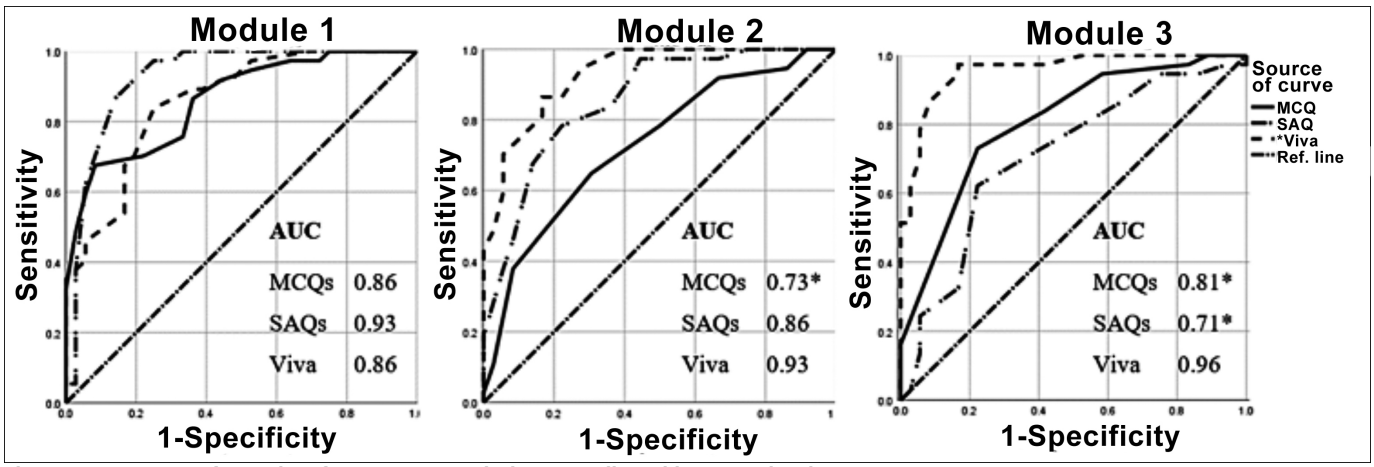

Figure 1: ROC curve of capacity of MCQs, SAQs and *Viva* to predict achievement level.

**Represents statistically significant difference of AUC from viva-voce. AUC = area under curve, MCQs = Multiple-choice questions, SAQs = Short answer questions*

In the second module exam, conducted online, *viva* was a much better discriminating item (0.47) as compared to the SAQs (0.24) and MCQs (0.20) that were marginal. The third module, also conducted online, had *viva* as an excellent discriminating part (0.49) of the exam in contrast to the MCQs (0.19) and SAQs (0.18) that showed poor discrimination power (Table I).

Table I: Discrimination index for MCQs, SAQs and *viva* for module exams.

Exam	MCQs	SAQs	*Viva*
Module 1	0.27	0.38	0.23
Module 2^{+}	0.20	0.24	0.47
Module 3^{+}	0.19	0.18	0.49

MCQs = Multiple-choice questions, SAQs = Short answer questions.
$^{+}$Represents the exam conducted online.

Table II: Area under curve (AUC) with 95% confidence interval for MCQs, SAQs and *viva* for three module exams.

	MCQs	SAQs	*Viva*
Module 1	0.86 (0.78 - 0.94)	0.93 (0.86 - 0.99)	0.86 (0.77 - 0.94)
Module 2^{+}	0.73 (0.61 - 0.84)	0.86 (0.77 - 0.94)	0.93 (0.87 - 0.98)
Module 3^{+}	0.81 (0.71 - 0.91)	0.71 (0.59 - 0.83)	0.96 (0.91 - 1.00)

MCQs = Multiple-choice questions, SAQs = Short answer questions.
$^{+}$Represents the exam conducted online.

Analysis of the area under the ROC curve showed that SAQs demonstrated outstanding discrimination between the high and low achievers in the first module exam with an AUC of 0.93. MCQs and *viva* also had excellent discrimination, and there were no statistically significant differences between the AUC of all three. In the second module exam, conducted online, the exam's *viva* component showed excellent discriminating ability, with AUC significantly higher than that of MCQs only (p = 0.003). The third module had *viva* with an AUC of 0.96, which was higher than the AUC of both MCQs (p = 0.007) and SAQs (p <0.001). AUC of MCQ was not different from that of SAQs (p=0.1, Table II, Figure 1).

DISCUSSION

This study aimed to compare the discrimination indices of each of the three exam components (MCQs, SAQs, and *viva-voce*) in on-campus and online examinations.

In the first module exam, conducted on-campus, the short answer questions (SAQs) had the highest discrimination index and AUC compared to the other components. AUC of MCQs and *viva* was less but not significantly different. This showed a relative superiority of SAQs in their ability to differentiate between the high and low scorers in the on-campus examination. These results were consistent with Dhakal *et al*.,[14] who found SAQs to have better discriminating ability than the MCQs in the second year MBBS anatomy exams. Similar results were found by Thomas *et al*.,[20] who found very short answer and questions (VSAQs) superior to multiple-choice questions (MCQs) after comparing test performance, difficulty and discrimination indices, and feedback by students and lecturers. Short answer questions' better discriminating ability can be explained by the fact that the SAQs employ directly asked questions that require students to reply briefly and directly with a little to no scope of guessing or test-wiseness.

Interestingly, this interplay was different for the exams conducted online. In both second and third module exams (conducted online), *viva-voce* had a much higher discrimination index and significantly higher AUC than the theory components, more so in the latter than the former exam. The more inferior discriminating ability of MCQs and SAQs can be attributed to several factors. In general, students are more comfortable and prefer online examinations as compared to paper exams.[6] They offer lower levels of during-test anxiety as compared to the traditional exams.[21] Last but not the least, there is an increased opportunity for

cheating during the remote exam. According to one survey, 73.6% of the students believe it is easier to cheat online than traditional exams.[22]

To quote Sullivan, "cheating, besides a 'crisis on-campus' and the 'most commonly reported challenge in online assessment' is 'reaching virtually pandemic proportions', and the expanding scale and scope of online education complicate circumstances."[23] The lack of specialised software, while conducting the exam, provided students with a window of opportunity to access help online; also, invigilation through webcam provided a limited field of view to exclude the presence of any helping material. In the authors' opinion, this was one of the principal factors in minimising the discriminating ability of the theory component of the examination, as discussed in one of their earlier works.[24]

On the other hand, the conditions of one-on-one *viva-voce* were very similar to that of in-person, live interaction giving little to no opportunity to get outside help. Even in on-campus examinations, the *viva-voce* is as effective or even better in evaluating students' understanding and application of knowledge than theory exams.[13,25] The factors mentioned above help explain why *viva-voce* became the chief factor in discriminating between the high and low scoring students in online examinations.

This study had a few limitations. It was a single-centre study, reporting results of only the Anatomy Department during one academic year. The study design was a naturalistic inquiry rather than a pre-designed cohort. These limitations must be kept in mind while interpreting and generalising these findings. Despite the limitations, this study is unique and pertinent in highlighting that *viva-voce* can serve a better discriminating component of online exams than on-campus exams in given resources. Thus, analysing and quantifying, online exams' standards can help policymakers invest their resources in deficient areas.

CONCLUSION

Viva-voce had a significantly higher discriminating index than multiple-choice questions and short answer questions in online exams. On-campus, SAQs had the highest discriminatory index. Using this statistical approach, the effectiveness of different components of the online exam can be monitored to improve the quality of online exams amidst the COVID-19 crisis.

ETHICAL APPROVAL:

The ethical approval of this study was obtained from the Institutional Review Board and Ethical Committee (IRB&EC) of CMH Multan Institute of Medical Sciences (Letter No. TW / 51 / CIMS).

PATIENTS' CONSENT:

Students' consents were not obtained as the study was designed on retrospective exam result analysis; and complete confidentiality and anonymity were ensured.

CONFLICT OF INTEREST:

The authors declared no conflict of interest.

AUTHORS' CONTRIBUTION:

MAR: Conceived idea, collection of data, manuscript writing.
HBSS: Interpretation of data, critical revision of the manuscript.
FI: Analysis of data, literature search, manuscript writing.

REFERENCES

1. Kakodkar P, Kaka N, Baig M. A comprehensive literature review on the clinical presentation, and management of the pandemic coronavirus disease 2019 (COVID-19). *Cureus* 2020; **12(4)**:7560. doi: 10.7759/cureus.7560.
2. Buheji M, da Costa Cunha K, Beka G, Mavrić B, Leandro do Carmo de Souza Y, Souza da Costa Silva S, *et al*. The extent of COVID-19 pandemic socio-economic impact on global poverty. A global integrative multidisciplinary review. *Am J Econ* 2020; **10(4)**:213-24.
3. Almarzooq ZI, Lopes M, Kochar A. Virtual learning during the COVID-19 pandemic: A disruptive technology in graduate medical education. *J Am Coll Cardiol* 2020; **75(20)**: 2635-8. doi: 10.1016/j.jacc.2020.04.015.
4. Konig J, Jager-Biela DJ, Glutsch N. Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany. *Eur J Teach Educ* 2020; **43(4)**: 608-22.
5. Mukhtar K, Javed K, Arooj M, Sethi A. Advantages, limitations and recommendations for online learning during covid-19 pandemic era. *Pakistan J Med Sci* 2020; **36(COVID19-S4)**:S27-31. doi: 10.12669/pjms.36.COVID19-S4.2785.
6. Butler-Henderson K, Crawford J. A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Comput Educ* 2020; **159(May)**:104024. doi: 10.1016/j.compedu.2020.104024.
7. Prigoff J, Hunter M, Nowygrod R. Medical student assess--ment in the time of COVID-19. *J Surg Educ* 2021; **78(2)**: 370-4. doi: 10.1016/j.jsurg.2020.07.040.
8. Jaap A, Dewar A, Duncan C, Fairhurst K, Hope D, Kluth D. Effect of remote online exam delivery on student experience and performance in applied knowledge tests. *BMC Med Educ* 2021; **21(1)**:86. doi: 10.1186/s12909- 021-02521-1.
9. Cramp J, Medlin JF, Lake P, Sharp C. Lessons learned from implementing remotely invigilated online exams. *J Univ Teach Learn Pract* 2019; **16(1)**. doi: 10.14453/jutlp.v16i1.10.
10. James R. Tertiary student attitudes to invigilated, online summative examinations. *Int J Educ Technol High Educ* 2016; **13(1)**:19.
11. Elsalem L, Al-Azzam N, Jum'ah AA, Obeidat N. Remote e-exams during Covid-19 pandemic: A cross-sectional study of students' preferences and academic dishonesty in faculties of medical sciences. *Ann Med Surg* 2021; **62**: 326-33. doi: 10.1016/j.amsu.2021.01.054.
12. Abdollahi A, Labbaf A, Mafinejad MK, Sotoudeh-Anvari M, Azmoudeh-Ardalan F. Online assessment for pathology residents during the COVID-19 pandemic: Report of an experi-

ence. *Iran J Pathol* 2020; **16(1)**:75-8. doi: 10.30699/ijp.2020.129558.2425.

13. Rushton P, Eggett D. Comparison of written and oral examinations in a Baccalaureate medical-surgical nursing course. *J Prof Nurs* 2003; **19(3)**:142-8. doi: 10.1016/ s8755-7223(03)00049-8.
14. Dhakal A, Yadav SK, Dhungana GP. Assessing multiple--choice questions (MCQs) and structured short-answer questions (SSAQs) in human anatomy to predict students' examination performance. *J Res Med Educ Ethics* 2018; **8(2)**:127.
15. Farooqui F, Saeed N, Aaraj S, Sami MA, Amir M. A comparison between written assessment methods: Multiple--choice and short answer questions in end-of-clerkship examinations for final year medical students. *Cureus* 2018; **10(12)**:e3773. doi: 10.7759/cureus.3773.
16. Miller DM, Linn RL. Measurement and assessment in teaching. 10th ed. Upper saddle river, NJ: *Pearson Education* 2009;
17. Kelley TL. The selection of upper and lower groups for the validation of test items. *J Educ Psychol* 1939; **30(1)**:17-24.
18. Jandaghi G, Shaterian F. Validity, reliability and difficulty indices for instructor-built exam questions. *J Appl Quant Methods* 2008; **3(2)**:151-5.
19. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression: Third edition. 3rd ed. Applied logistic regression: Third edition. Hoboken, NJ: John Wiley & Sons, Inc; 2013.
20. Puthiaparampil T, Rahman MM. Very short answer questions: A viable alternative to multiple choice questions. *BMC Med Educ* 2020; **20(1)**:1-8. doi: 10.1186/s12909-020-02057-w.
21. Stowell J, Bennett D. Effects of online testing on student exam performance and test anxiety. *J Educ Comput Res* 2010; **42(2)**:161-71.
22. Aisyah S, Bandung Y, Subekti LB. Development of continuous authentication system on android-based online exam application. In: 2018 International conference on information technology systems and innovation, ICITSI 2018 - proceedings. Institute of electrical and electronics engineers Inc.; 2018. p. 171-6.
23. Sullivan DP. An integrated approach to preempt cheating on asynchronous, objective, online assessments in graduate business classes. *Online Learn J* 2016; **20(3)**: 195-209.
24. Ikram F, Rabbani MA. Academic integrity in traditional *vs* online undergraduate medical education Amidst COVID-19 Pandemic. *Cureus* 2021; **13(3)**. doi: 10.7759/cureus.13911.
25. Huxham M, Campbell F, Westwood J. Oral *versus* written assessments: A test of student performance and attitudes. *Assess Eval High Educ* 2012; **37(1)**:125-36. doi: 10.3205/zma001170.