ORIGINAL ARTICLE

Viva Voce Examination Using Unstructured Impromptu Questions and Structured *Viva Voce* Cards: A Comparative Study among Final Year MBBS Students in the Subject of Internal Medicine

Abraham M Ittyachen[1*], Binitha Baby[2], Meera B John[3], Neha M Baby[4], Sarath C Mathew[5]

Received: 06 January 2025; *Accepted:* 23 May 2025

Abstract

Background: In the medical curriculum, *viva voce* is a crucial component of formative and summative assessment. However, despite many advantages, it also has some pitfalls. The primary objective of this study is to ascertain whether there is a significant difference in assessment between *viva voce* conducted using extempore questions and *viva voce* conducted using structured *viva voce* cards.

Methods: This prospective observational study took place over a period of 3 months. *Viva voce* was conducted by four examiners in internal medicine among final-year students pursuing MBBS, initially in an unstructured manner using extempore questions, and later using structured questions of increasing difficulty in the form of printed cards. A theory examination was conducted before the *viva voce*. In addition, a feedback survey using a Likert scale questionnaire was conducted among the students and examiners to assess their perception.

Results: Students scored the best in unstructured *viva*, followed by theory and structured *viva*. There was a moderately positive correlation between unstructured and structured *viva* and theory scores. There was a poor correlation between *viva* and perception scores for both unstructured and structured *viva*. The examiners had a slightly more positive perception toward the unstructured *viva* method than the structured method.

Conclusion: Unlike most studies in preclinical and paraclinical subjects, this study is a contradiction, with students scoring better in the unstructured method and faculty also showing a preference for the same. Further research is required on the effectiveness of both unstructured and structured *viva* in clinical settings.

Journal of The Association of Physicians of India (2025): 10.59556/japi.73.1083

Introduction

"*Viva Voce*" (Medieval Latin) or examination "by word of mouth," "orally," and "by the living voice," was the earliest form of formal assessment of medical and other apprentices, dating back to premedieval times.[1] Currently also, *viva voce* forms an important part of formative and summative evaluation in medical courses. It allows the examiner to test the scope of knowledge of the subject[2] and probe the limits of knowledge in both borderline and exceptional students.[3] However, the method has been an area of contention because of excessive subjectivity and being swayed by academic and nonacademic factors associated with students and teachers.[3] Low validity and low reliability are the other disadvantages of the traditional *viva*.[4]

Research Question

Is there a significant difference in the assessment between *viva voce* conducted using extempore questions and in a structured manner using *viva voce* cards?

Objectives

- To estimate whether there was a significant difference in assessment between *viva voce* conducted using extempore questions and in a structured manner using *viva voce* cards.
- To compare the marks, the examinee (student) obtained in the two *viva voce* methods with those obtained in the traditional theory paper.
- To conduct a postsurvey study using a Likert scale questionnaire among students and examiners to assess their perception of the study.

Review of Literature

Viva voce, or conventional oral examination, is an important instrument for evaluation in medical education. However, common drawbacks are that it is affected by both academic and nonacademic factors that pertain to students and teachers, which led to the criticism for being too subjective.[3]

The test environment and the candidate's level of anxiety may affect scores in the traditional form of *viva voce*.[5] The examination scores were also correlated with personality scores.[6] The system-related factors described are central tendency, leniency, "Halo effect," and error of contrast.[7]

One important drawback of the conventional oral examination is examiner "fatigue." Students who are initially examined may be asked many questions, but as the examination proceeds, students toward the end tend to be asked fewer questions for the same duration. This contributes to subjective variation in the assessment of candidates. In addition to the high subjectivity of the examiner,[3,8–10] *viva voce* may also be taken casually. Studies have shown that scores are directly proportional to the number of words spoken by the examiners and the time taken by them.[11]

Overall, conventional oral examinations are more time consuming and much less cost-effective than other methods of examination.[12,13]

Examiners may be prejudiced by their knowledge of the subject, their choice areas of interest, and momentary environmental distractions.[14–16] The standard of the questions asked, the self-esteem of the examiner, and the order in which the questions are asked, sometimes the students felt dissatisfied and disgraced. In some cases, the event is felt as intimidating and threatening, which at times gives way to confrontation.[15–17]

[1]Professor of Medicine and Consultant in Critical Care; [2–5]Assistant Professor, Department of General Medicine, Malankara Orthodox Syrian Church Medical College Hospital, Ernakulam, Kerala, India; *Corresponding Author

How to cite this article: Ittyachen AM, Baby B, John MB, *et al. Viva Voce* Examination Using Unstructured Impromptu Questions and Structured *Viva Voce* Cards: A Comparative Study among Final Year MBBS Students in the Subject of Internal Medicine. J Assoc Physicians India 2025;73(8):50–54.

Examiners may discriminate on the basis of economic status, gender, ethnic status, or minority status. They may even be influenced by personality, clothing, and verbal style of the candidate, in *viva voce*.[15]

Methodology

This prospective observational study (cohort study) was conducted in the Department of Medicine of Malankara Orthodox Syrian Church Medical College Hospital, Ernakulam, Kerala, India. The study period was 3 months (from April 1 to June 30, 2024), and the study subjects were final-year MBBS students.

Inclusion Criteria

All final-year MBBS students (2020 batch) participated in the second sessional exam and were ready to participate in the study.

Exclusion Criteria

Final-year MBBS students (2020 batch) participated in the second sessional exam and were not willing to be part of the study.

After securing informed consent, the study objectives were explained to the participating students. On the 1st day, four examiners assessed the students using *viva voce* in an unstructured manner by asking questions; each examiner was restricted to one subject. The next day, the same four examiners asked four questions using structured cards, each of graded difficulty (one to four marks, one for the easiest question among the four, and four for the most difficult), and marks were awarded based on the number of questions answered correctly. Each examiner was restricted to one participant. The questions were arranged in a manner that progressively increased in difficulty, in accordance with Bloom's taxonomy of educational objectives of the cognitive domain, and were collectively endorsed by the examiners. Theory examination was conducted before the *viva voce*. The topics covered in the theory and *viva* were the same. Care was taken to ensure that questions on the printed cards were not asked during the unstructured *viva voce*. In addition, a feedback survey using a Likert scale questionnaire was conducted among the students and examiners to assess their perception toward the two methods of *viva* (Fig. 1). A questionnaire was developed specifically for this study. The study protocol was approved by the Institutional Ethics Committee (IEC): (MOSC/IEC/127/2024).

Statistical Analysis

Descriptive statistics were used to detail students' performance across the two examination formats. The Wilcoxon signed-rank test was conducted to compare structured and unstructured *viva* scores, and in the same way for structured and unstructured *viva* scores with theory scores. The association between student gender, with two defined categories, and examination scores was also explored. If the data within each sex category followed a normal distribution, independent *t*-tests were used to compare the mean scores between male and female. For nonnormally distributed data, the Mann–Whitney *U* test was used. Correlation analysis was used to examine the relationship between continuous variables, such as the marks obtained in the two examination formats, and between marks and sex. The Pearson correlation coefficient was computed for data that followed a normal distribution, while Spearman's rank correlation was applied to data that did not conform to normality. A statistical significance threshold of $p < 0.05$ was established. The statistical analyses were conducted utilizing IBM Statistical Package for the Social Sciences (SPSS) version 22.

The rationale behind testing the correlation between perception and *viva* scores was to ascertain if a statistical relationship existed between how students perceived their understanding of the subject and their actual performance in *viva* (oral examination). A strong correlation suggests that student's perceptions accurately reflect their knowledge, whereas a weak correlation indicates a mismatch between perception and actual performance.

Results

Collectively, 83 students were part of the study, of whom 15 (18.1%) were male and 68 (81.9%) were female. Average age of the study participants was 22.5 ± 0.75 (min: 21, max: 24).

A statistically significant difference was observed in the average scores of students (subject-wise) when comparing unstructured and structured *viva* voices, with students scoring better in unstructured *viva* (Table 1).

When comparisons were made between unstructured *viva* scores (subject-wise) and theory marks, a statistically significant difference in the scores of students was noted, with students scoring better in unstructured *viva* (Table 2A). However, a comparison between the structured *viva* scores (subject-wise) and theory indicated that students scored better in theory (Table 2B).

When the subjects were taken together, there was a statistically significant difference in the average score of the students, with students scoring better in the unstructured *viva* than structured *viva* (Fig. 2).

Spearman rank correlation showed a moderate positive correlation between structured *viva* scores and theory scores ($r = 0.483$, p-value < 0.001) and unstructured viva scores with theory scores ($r = 0.496$, p-value < 0.001).

Among males, there was a moderate positive correlation between the structured *viva* and theory scores, but this did not

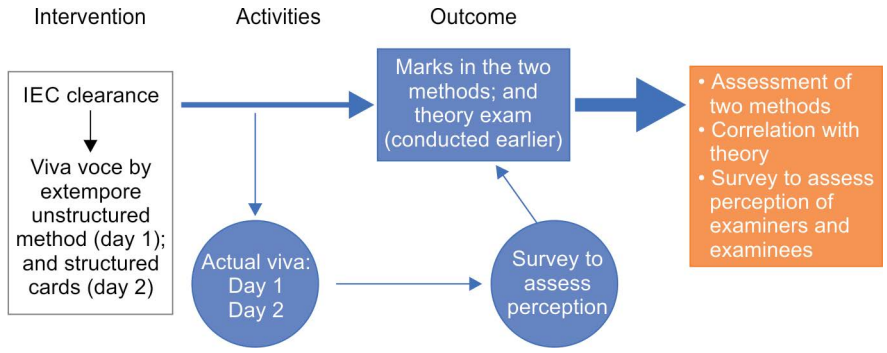

Fig. 1: Schematic representation of the research methodology

Table 1: Comparison of unstructured *viva* scores with structured *viva* scores

Subjects	*Unstructured viva score median (Q1, Q3)*	*Structured viva score median (Q1, Q3)*	*Test statistic value*	*p-value*
Cardiovascular system (CVS)	4.5 (4.0, 5.0)	2.0 (1.0, 3.5)	6.73	<0.001*
Respiratory system (RS)	6.0 (4.0, 7.0)	2.0 (1.5, 3.0)	7.73	<0.001*
Nephrology	5.0 (5.0, 6.0)	3.0 (1.0, 4.0)	7.44	<0.001*
Gastrointestinal tract (GIT)	6.0 (5.0, 7.0)	2.0 (1.0, 4.0)	7.67	<0.001*

*Statistically significant at $p < 0.05$

Table 2: Comparison of *viva* scores (subject-wise) with theory scores

Subjects	*Unstructured viva score median (Q1, Q3)*	*Theory marks median (Q1, Q3)*	*Test statistic value*	*p-value*
(A) Unstructured				
Cardiovascular system (CVS)	4.5 (4.0, 5.0)	3.5 (2.6, 4.5)	5.42	<0.001*
Respiratory system (RS)	6.0 (4.0, 7.0)	3.5 (2.6, 4.5)	7.42	<0.001*
Nephrology	5.0 (5.0, 6.0)	3.5 (2.6, 4.5)	6.90	<0.001*
Gastrointestinal tract (GIT)	6.0 (5.0, 7.0)	3.5 (2.6, 4.5)	7.53	<0.001*
Subjects	*Structured viva score median (Q1, Q3)*	*Theory marks median (Q1, Q3)*	*Test statistic value*	*p-value*
(B) Structured				
CVS	2.0 (1.0, 3.5)	3.5 (2.6, 4.5)	4.78	<0.001*
RS	2.0 (1.5, 3.0)	3.5 (2.6, 4.5)	5.40	<0.001*
Nephrology	3.0 (1.0, 4.0)	3.5 (2.6, 4.5)	4.29	<0.001*
GIT	2.0 (1.0, 4.0)	3.5 (2.6, 4.5)	4.89	<0.001*

*Statistically significant at $p < 0.05$

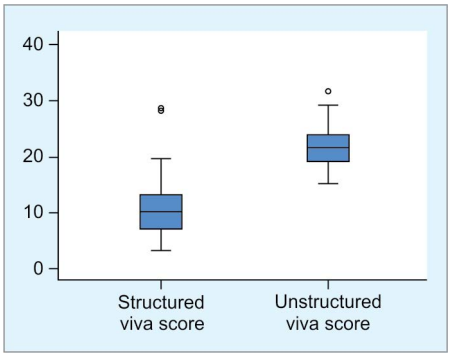

Fig. 2: Comparison of total score (total of all subjects) of structured and unstructured *viva*

reach statistical significance ($r = 0.401$, p-value = 0.139). Conversely, a moderate positive correlation was identified between the unstructured *viva* and theory scores which was statistically significant ($r = 0.604$, p-value = 0.017). Among females, a statistically significant moderate positive correlation existed between the structured *viva* scores ($r = 0.467$, p-value < 0.001), unstructured *viva* scores ($r = 0.487$, p-value < 0.001), and theory scores.

The Mann–Whitney *U* test revealed no statistically significant difference in structured [11.5 (6, 12.5) vs 10 (7, 13), p-value = 0.873] and unstructured *viva* scores [21.50 (18, 23) vs 21.25 (19, 24), p-value = 0.785] between males and females.

In the questionnaire to assess the perception of students with regards to the structured *viva voce* method, there were eight questions rated on a Likert scale, with a lowest possible score of eight and a highest possible score of 40. The median score was 27 (23, 29). Questionnaire reliability was assessed using Cronbach's alpha. The alpha value was 0.760, which is good and acceptable. The Spearman's correlation coefficient revealed a poor correlation ($r = 0.017$) between the structured *viva voce* and perception scores. The Mann–Whitney *U* test was used to compare the perception score between genders, with a median (Q1, Q3) score of 27 (23, 29.75) for females and 26 (26, 28) for males, which was statistically insignificant (p-value = 0.565).

In the questionnaire used to assess the perception of students toward the unstructured *viva voce* method, there were six questions on a Likert scale, with a minimum score of six and a maximum score of 30. The median score was 21 (19, 23). Questionnaire reliability was assessed using Cronbach's alpha. The alpha value was 0.639, which was acceptable. The Spearman's correlation coefficient revealed a poor correlation ($r = 0.181$) between the unstructured *viva voce* and perception scores. The Mann–Whitney *U* test was used to compare the perception scores between genders, with median (Q1, Q3) scores of 21 (19, 23) for females and 21 (20, 23) for males, which was not statistically significant (p-value = 0.374).

The feedback survey using a Likert scale among the four examiners to assess their perceptions toward structured *viva voce* and unstructured *viva voce* methods consisted of 10 questions, with a minimum score of 10 and a maximum score of 50. The median perception score of examiners regarding their structured *viva voce* method was 31 (30, 33.5), whereas for unstructured *viva voce* method it was 33.5 (30.75, 34.75).

Discussion

The *viva voce* method of examination forms an important part of assessment in the medical education. In this study, an attempt was made to determine whether there is a significant difference in the assessment between *viva voce* conducted using extempore questions and *viva voce* conducted in a structured manner using *viva voce* cards in the subject of internal medicine. Most studies available in the literature have compared unstructured and structured *viva voce* examinations involving pre- and paraclinical subjects. This study involved a clinical discipline (internal medicine) and, therefore, may be considered rare.

When comparison was made in this study between unstructured and structured *viva voce*, a statistically significant difference in the average score of students was noted, with students scoring better in unstructured *viva*. This was true when the analysis was conducted subject-wise and when done with the subjects taken together. In a study by Khilnani et al.,[2] *viva voce* was conducted with both conventional and structured methods among undergraduate students in pharmacology; structured *viva* yielded significantly lower marks compared to conventional *viva*. However, this was not true for all subjects. When a similar study was conducted among undergraduate anatomy students, no major difference was found in mean scores between the two methods.[17] Across specialties, when the academic performance of undergraduate dental students was compared using structured and unstructured oral examinations, there was no major difference in mean scores achieved.[18] Since no single method, structured or unstructured, can be considered superior across the spectrum of undergraduate medical subjects, further studies should be conducted on this topic. In a systematic review and meta-analysis by Anbarasi et al.,[19] 18 peer-reviewed articles on conventional and structured oral examinations for medical students were reviewed. The analysis indicated that there was no difference in the mean marks obtained by the conventional *viva* or structured method.

The cognitive domain is the most important of all domains. Psychomotor and affective domains are inextricably linked to the cognitive domain. Without appropriate intelligence/knowledge of the subject, it is unlikely that the student will demonstrate exceptional skills or conduct. In medical courses, the cognitive domain is primarily expressed in the theory paper.[20]

In this study, when comparing unstructured *viva* scores (subject-wise) and theory marks, there was a statistically significant difference in the scores of students, with students scoring better on the unstructured *viva*. If theory is considered the gold standard in the assessment of the cognitive domain, a higher score in unstructured *viva* indicates a poor correlation between the two, and unstructured *viva* is a poor marker of the cognitive domain.

When comparison was made between structured *viva* scores (subject-wise) and theory marks in this study, there was a statistically significant difference in the scores of students, with students scoring better in theory. Studies have shown a better correlation between theory and structured *viva* than unstructured *viva*.[21,22] A properly designed structured *viva* has multiple advantages in that it has wide coverage; it promotes disinhibition, encourages better expression, and reduces anxiety and shyness in the student, in contrast to an unstructured *viva* that is stained with high subjectivity, lack of a format and uniformity,[9] and unreliability.[15,23] An unstructured *viva* also heightens apprehension among students (strenuous level of questions, emphasis on problem-solving, and direct and immediate feedback) and hesitation among faculty members (structured *viva* demands comprehensive planning, prevalidated well-structured questions, scoring criteria, and adequate resources and manpower) in terms of execution.[24]

The Spearman rank correlation showed a moderate positive correlation between unstructured and structured *viva* and theory scores in this study. However, when comparisons were made (as explained above) between the scores obtained by the two *viva* methods (unstructured and structured) and theory, there was a statistically significant difference in the score with students scoring better in unstructured *viva*. However, this was not the case for structured with students scoring better in theory.

There was no statistically significant difference in structured and unstructured *viva* scores between males and females in this study. However, studies involving the traditional (unstructured) *viva voce* have highlighted the presence of possible gender bias. Lack of standardization leading to variability in questioning and assessment can inadvertently introduce gender bias. By implementing a standardized set of questions and evaluation criteria, structured *viva* can provide a uniform platform for all candidates thereby reducing the scope to gender bias.

When the questionnaire assessing the perception of students toward unstructured *viva voce* was analyzed, there was a poor correlation between *viva* and perception scores. So was the conclusion with structured *viva*.

When the feedback survey from the examiners to assess their perception toward the two types of *viva* was analyzed, the median scores and the ranges provided suggested that the examiners had a slightly more positive perception toward the unstructured *viva* method. This information could be useful in understanding the examiner's preferences and perceptions toward the two different *viva voce* methods, which could in turn decide the choice of assessment method or the need for further studies.

Conclusion

Students scored best in unstructured *viva voce* setting, followed by theory and structured *viva*. There was a moderately positive correlation between unstructured and structured *viva* and theory scores. There was a poor correlation between the *viva* scores and the perception (toward *viva*) scores of students. Examiners had a slightly more positive perception toward the unstructured *viva* method. Most of the studies showing the acceptability of structured *viva* have dealt with pre- and paraclinical subjects, whereas studies on clinical subjects in which the two *viva* methods are compared are rare. Unlike studies involving pre- and paraclinical subjects that have shown acceptability for structured *viva*, this study revealed an opposite trend, with students scoring better in the unstructured method and faculty also showing a preference for the same. Further research is required to assess the efficacy of both unstructured and structured *viva voce* examination in clinical subjects to assess learning progress.

Another option would be to modify the traditional *viva voce* methods to bring out the best of the two methods studied here. Combining the structured and unstructured elements (hybrid approach) can offer a balanced assessment. For instance, starting with standardized questions to ensure core competencies are covered, followed by open-ended questions that allow exploration of the student's clinical reasoning and problem-solving abilities.

Limitations

This study has some limitations. The sample size was small because only students from one institution participated in the study. Moreover, this was a single-center study and may not represent a wider population of medical students. This study involved a clinical discipline and, as already mentioned, similar studies involving a clinical subject are rare; hence, further research involving other clinical disciplines is needed. Additionally, methodological limitations include variables that could affect student's performance, such as prior knowledge or test anxiety, which may not have been controlled for. The presence of potential confounders, and effect modifiers is also a limitation.

As a study this could have potential confounders and effect modifiers. First, the examiner bias where the results of the assessment may be affected by the prejudgments or preferences of individual examiners for particular *viva voce* techniques. Experienced examiners might handle impromptu questioning more proficiently, while less experienced ones might rely heavily on structured formats. This discrepancy can affect the consistency and fairness of assessments. Second, student's readiness representing differences in student's preparedness or familiarity with the various *viva voce* forms may have a direct bearing on the outcome. For instance, those familiar with structured questions might excel in that format but struggle with impromptu questions, leading to performance differences unrelated to actual knowledge or skills.

Prolonged assessment periods can lead to examiner fatigue, potentially affecting their concentration and judgment. Examiners' personal biases and personality traits can introduce variability in student's assessments. The interplay of examiner fatigue, biases, and personality can create an inconsistent assessment environment, potentially disadvantaging students.

Finally, the examination room environment, such as comfort, noise level, distractions, and overall atmosphere can influence student's outcomes.

Acknowledgments

The authors wish to express their gratitude to the students of the 2020 MBBS batch of Malankara Orthodox Syrian Church Medical College Hospital, Ernakulam, who participated in the study; the staff of the Department of Medicine, Malankara Orthodox Syrian Church Medical College Hospital, Ernakulam; Mrs Anjaly S Nair, lecturer and biostatistician in Malankara Orthodox Syrian Church Medical College Hospital, Ernakulam; and the Faculty of National Medical Commission (NMC) Nodal Center for Faculty Development, Government Medical College, Kottayam, India.

Orcid

Abraham M Ittyachen https://orcid.org/0000-0001-7918-5500
Neha M Baby https://orcid.org/0009-0000-1392-0277

References

1. Bode C, Ugwu B, Donkor P. *Viva voce* in postgraduate surgical examinations in anglophone West Africa. J West Afr Coll Surg 2011;1(1):40–52.
2. Khilnani AK, Charan J, Thaddanee R, et al. Structured oral examination in pharmacology for undergraduate medical students: factors influencing its implementation. Indian J Pharmacol 2015;47(5):546–550.
3. Torke S, Abraham R, Komattil R, et al. The impact of *viva voce* examination on students' performance in theory component of final summative evaluation in physiology. J Physiol Pathophysiol 2010;1:10–12.
4. Abuzied AIH, Nabag WOM. Structured viva validity, reliability, and acceptability as an assessment tool in health professions education: a systematic review and meta-analysis. BMC Med Educ 2023;23(1):531.
5. Iqbal IZ, Naqvi S, Abeysundara L, et al. The value of oral assessments: a review. Ann R Coll Surg Engl 2010;92:1–6.
6. Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. Adv Health Sci Educ Theory Pract 2010;15(2):277–289.
7. Evans LR, Ingersoll RW, Smith EJ. The reliability, validity, and taxonomic structure of the oral examination. J Med Educ 1966;41(7):651–657.
8. Ray MK, Ray S, Ray U. Technology enabled assessment of *viva voce*: a new challenge. J Adv Res Biol Sci 2013;5:238–242.
9. Ghosh A, Mandal A, Das N, et al. Student's performance in written and viva-voce components of final summative pharmacology examination in MBBS curriculum: a critical insight. Indian J Pharmacol 2012;44(2):274–275.
10. Verma A, Mahajan N, Jasani K. Evaluation & comparison of results: conventional viva vs. structured viva. Glob Res Anal 2013;2(5):188–189.
11. Holloway PJ, Hardwick JL, Morris J, et al. The validity of essay and viva-voce examining techniques. Br Dent J 1967;123(5):227–232.
12. Oakley B, Hencken C. Oral examination assessment practices: effectiveness and change with a first year undergraduate cohort. J Hosp Leis Sport Tourism Educ 2005;4(1):3–14.
13. Schuwirth LWT, van der Vleuten CPM. Changing education, changing assessment, changing research? Med Educ 2004;38(8):805–812.
14. Singel TC, Shah C, Dixit D. Small group structured oral examination: an innovation in oral examination. Natl J Integr Res Med 2014;5(1):141–145.
15. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. Med Teach 2005;27(4):294–297.
16. Haque M, Yousuf R, Abu Bakar SM, et al. Assessment in undergraduate medical education: Bangladesh perspectives. Bangladesh J Med Sci 2013;12(4):357–363.
17. Shaikh ST. Objective structured viva examination versus traditional viva examination of medical students. Anat Physiol 2015;5(3):175.
18. Lakshminarayan N, Usha GV. Comparison of undergraduate dental students' academic performance using structured and unstructured oral examinations. Int J Acad Med 2022;8(3):131–136.
19. Anbarasi K, Karunakaran JV, Ravichandran L, et al. Effectiveness of the structured and conventional methods of viva examination in medical education: a systematic review and meta-analysis. J Clin Diagn Res 2022;16(9):JE01–JE07.
20. Shukla S, Acharya S, Acharya N, et al. Quality theory paper writing for medical examinations. J Clin Diagn Res 2014;8(4):XC01–XC04.
21. Dhasmana DC, Bala S, Sharma R, et al. Introducing structured *viva voce* examination in medical undergraduate pharmacology: a pilot study. Indian J Pharmacol 2016;48(Suppl 1):S52–S56.
22. Ahsan M, Mallick AK. A study to assess the reliability of structured viva examination over traditional viva examination among 2nd-year pharmacology students. J Datta Meghe Inst Med Sci Univ 2022;17(3):589–594.
23. Muzzin LR, Hart L. Oral examinations. In: Neufeld VR, Norman GR, editors. Assessing Clinical Competence. New York: Springer Publishing Company; 1985. pp. 71–93.
24. Rahman G. Appropriateness of using oral examination as an assessment method in medical or dental education. J Educ Ethics Dent 2011;1(2):46.