

Ann Natl Acad Med Sci (India), **52(3)**: 173-182, 2016

Short Answer Open-Ended versus Multiple-Choice Questions: A Comparison of Objectivity

Bharati Mehta¹, Bharti Bhandari¹, Parul Sharma², Rimplejeet Kaur³

¹Department of Physiology, All India Institute of Medical Sciences, Jodhpur
²RAK College of Medical Sciences, UAE; ³Dr. S. N. Medical College, Jodhpur.

ABSTRACT

Objectives: We designed our study with the hypothesis that open ended Short Answer type Questions (SAQs), no matter how carefully framed, cannot be as objective as Multiple Choice type Questions (MCQs).

Methods: The study was conducted on 1st year MBBS students (n=99) studying at AIIMS, Jodhpur. A written test on 'Blood & Immunity' was conducted containing same questions in two formats; twelve MCQs (type E) in section A and 12 SAQs in section B. Maximum marks for all questions in both sections were equal. All the answers of section B were evaluated separately by two different examiners to reduce the subjectivity and a model answer sheet for both the sections was prepared and provided to both the examiners.

Results: The difference in the scores in Section B SAQs that were evaluated by two different examiners was not statistically significant. Mean of the marks awarded by the two examiners was taken as the final score of each student in section B. The difference in the scores by the students in the two sections was also non-significant (p=0.14). A significant correlation (r=0.99, p<0.0001) was found in SAQ and MCQ scores. Bland-Altman analysis also showed no proportion of bias and the two methods of scoring were in agreement with each other.

Conclusion: The results suggest that meticulously-framed open-ended short answer type questions can be as objective as multiple choice type questions.

Keywords: Multiple choice questions, medical education, assessment, open-ended questions.

Correspondence : Dr. Bharti Bhandari, Assistant Professor, Department of Physiology, All India Institute of Medical Sciences, Jodhpur, Rajasthan, India. Mob. no. +91 8003996865, E-mail : drbhartibhandari@yahoo.co.in.

INTRODUCTION

Over the last few decades in India, there has been substantial re-evaluation in the undergraduate medical curriculum, especially on the teaching and assessment methodology. Assessment is a means of measuring knowledge and competence, has a pivotal role in stimulating learning as well as providing feedback to students and teachers(1). Written assessment is an integral part of medical education. Subjective assessment in theory is gradually giving way to objectivity; longer essay-type questions (LAQs) are being replaced by Short Answer Questions (SAQs) and Multiple Choice Questions (MCQs) (2). There have been much debate with respect to the type of written assessment to be administered to the students in order to test higher order cognition(3).

An assessment has to be reliable and valid, free from bias and manipulations (1, 2). Reliability is the degree to which an assessment tool produces stable and consistent results. Validity refers to how well a test measures what it is purported to measure (4). Open-ended format in the form of long-essay type, mini-essay type and short answer type questions is still preferred format for summative assessment. It is well documented that open ended questions have greater validity and they test higher order cognition level of knowledge (3, 5,6). On the other hand, closed-ended or

multiple choice format is shown to be reliable and efficient, objective, unbiased and make linguistic skills redundant. However MCQs can provide the unprepared students an opportunity to score if they guess right. One of the disadvantages of open ended questions is its low objectivity and reliability (7, 8).

Physiology as a discipline, creates a framework for understanding the normal functioning of the human body. It is a concept based subject and therefore has a scope for testing higher order cognitive skills with conceptual questions which, though open ended, have a precise answer. We designed our study with the hypothesis that open ended SAQs, no matter how carefully framed, cannot be as objective as MCQs.

Methods:

This Cross-sectional study was conducted on 1st year MBBS students (n=99) of All India Institute of Medical Sciences, Jodhpur in the department of Physiology, in 2014. A class test from 'Blood and Immune System' was declared about 2 weeks prior to the commencement of the study.

The following format of the test was decided. The students were not told about the type/format of the questions (Fig.1).

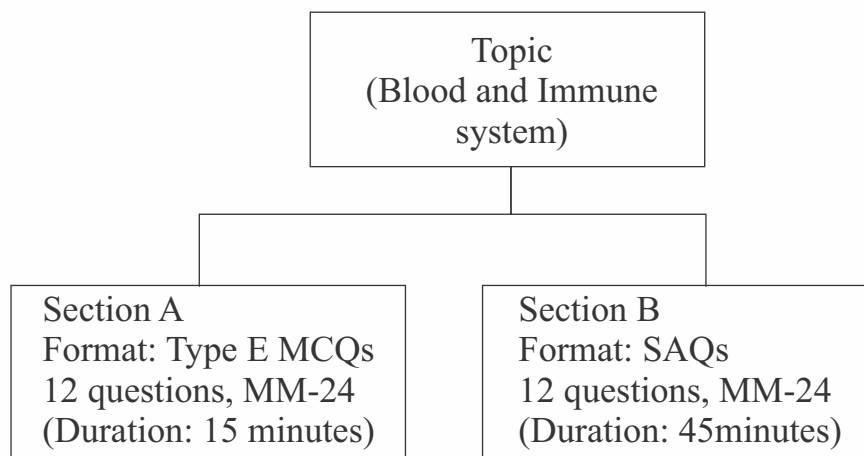


Fig. 1: Algorithm of topic and questions in Section A and B.

Both the sections comprised same questions but in different formats; MCQs and SAQs. Nevertheless, the questions were prepared by two different examiners. Marks for all questions in both sections were equal. Section B paper was distributed after collecting section A answer sheets. All the answers of section B were evaluated separately by two different examiners to minimize the level of subjectivity. A model answer sheet for both sections was prepared and provided to both the examiners. One of the examiners was the one who had taught the topic to these students during their routine teaching sessions.

There was no negative marking in

either section of the question paper.

Table 1 shows the questions in both the formats.

Section A consisted of Assertion-Reason (Type-E) multiple choice questions with following 5 options.

- A. Both A and R are true and R is the correct explanation of A.
- B. Both A and R are true but R is NOT the correct explanation of A.
- C. A is true but R is false.
- D. A is false but R is true.
- E. Both A and R are false.

Section B composed same concept based questions in SAQ format (explain why?)

Table 1 :Questions on identical subtopics in both the formats

Q.No.	Section A (MCQ format) Assertion-Reason	Section B (SAQ format) Explain why?
1	A- Globulin is the first plasma protein to appear in urine in renal diseases. R- Globulins are the smallest protein molecules in blood.	Albumin is the first plasma protein to appear in urine in renal diseases.
2	A- Jaundice is more common in the neonate than in the fetus. R- The fetus has a higher capacity to conjugate bilirubin than the neonate.	Jaundice appears in neonates but not in fetus.
3	A- ABO incompatibility in mother and fetus causes erythroblastosis fetalis. R- Fetal IgG can cross the placental barrier.	ABO incompatibility does not occur in mother and fetus.
4	A- Oxalate is the preferred anticoagulant during dialysis. R- Oxalate is metabolised in the body through the Krebs cycle.	Heparin is the preferred anticoagulant for dialysis.
5	A- Hypocalcemia impairs blood clotting. R- Calcium ion is essential for blood coagulation.	Persons with hypocalcemia never show clotting abnormalities.
6	A- The bleeding time is prolonged in obstructive jaundice. R- Obstructive jaundice is associated with poor absorption of vitamin K.	Clotting time but not bleeding time is prolonged in Obstructive jaundice.
7	A- Hypoproteinemia is associated with edema. R- Significant amount of plasma proteins are lost in the exudate.	Hypoproteinemia leads to generalised edema.
8	A- The hemolytic disease of the new- born is severe when a “B negative” mother bears an A positive fetus in the previous pregnancy. R- A “B negative” mother bearing an A positive fetus produces anti- B antibodies in the late third trimester.	The hemolytic disease of the new-born is less severe if the mother is B- and the previous baby was A+.
9	A-The secondary immune response is rapid and pronounced. R-B and T lymphocytes undergo blast transformation when exposed to antigens.	The secondary immune response is rapid and pronounced.
10	A-The mother well tolerates the fetus. R-The mother and fetus have same genetic makeup.	Fetus is a ‘transplant’ in the mother, yet it is well tolerated.
11	A- Decrease in helper T-cells decreases humoral immunity. R- Helper T- cells are essential for Tc- cell activity.	Decrease in helper T cells (as in AIDS) decreases, not only cellular, but also humoral immunity.
12	A- Immunity normally does not develop against ‘self’ antigens. R- Specific immunosuppression abolishes the response to ‘self’ antigens.	Immunity normally does not develop against self-antigens.

Question on identical subtopics in both the formats :

Statistical analysis was done using statistical packages; SPSS version 21 and GraphPad Prism version 6. Data on students' performance is expressed in percent Mean \pm SD. Unpaired t-test was applied to check inter-examiner bias. The mean and standard deviation, median and interquartile range of marks was calculated and scores in the 2 methods were compared using Wilcoxon Signed Rank test. Regression analysis was performed to see the association of the two scores. Spearman's correlation coefficient (r) was calculated to see the correlation between students' performance on short answer questions and the MCQs. Bland Altman analysis of differences and averages between SAQ and MCQ scores

was done (9). Mean bias and limits of agreement (mean bias \pm 1.96 times SD) were computed.

Results:

Mean marks awarded by examiner-1 and examiner-2 in all 12 questions in section B (open ended questions) were 14.44 ± 2.61 and 14.69 ± 2.82 , respectively, the difference was not statistically significant. Mean of the marks awarded by the two examiners was taken as the final score of each student in section B. **Table 2** depicts the percent marks (mean & median) scored by the students in section A and B. The difference in the scores was non-significant ($p=0.14$). A significant correlation ($r=0.99$, $p<0.0001$) was found in SAQ and MCQ scores (Fig.2, Table 2).

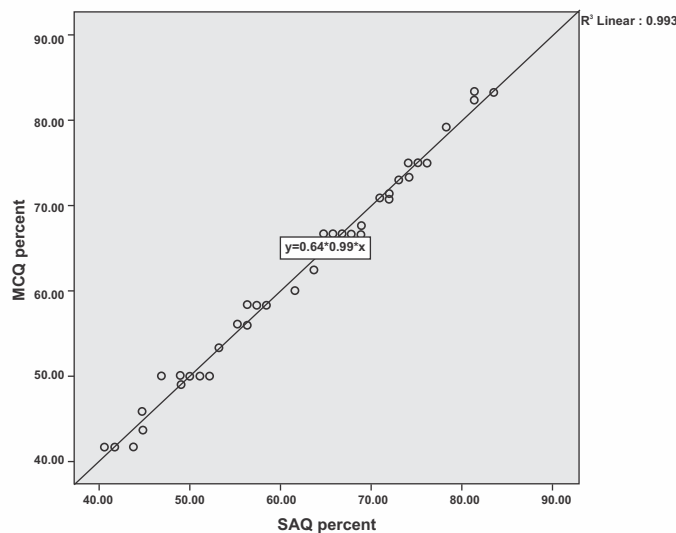


Fig. 2 : Scatter plot of the marks obtained in MCQ versus SAQ. Also shown is the linear best-fit line (Spearman $r = 0.99$, $p<0.0001$).

Table 2 : Marks awarded to students in section A and B.

No. of students	Section A - MCQ score (%)	Section B -SAQ score (%)	MCQ Vs SAQ score	
99	Mean: 60.92 ±11.23 (95%CI-58.68-63.16) Median: 41.67 (33.33-50)	Mean: 60.74 ±11.28 (95% CI-58.49-62.99) Median: 42.7(31.3-55.2)	Wilcoxon Signed Rank Test p=0.14 (ns)	Spearman's correlation r=0.99, p<0.0001

There was no significant difference in the marks scored by the students in the 2 sections. Significant correlation was observed in the 2 scores.

it is showing considerable agreement between the two methods of assessment (non-significant mean difference of 0.058), regression test showed non-significant results suggesting no proportion of bias ($t=-0.48, p=0.63$).

Figure 3 shows the Bland Altman Plot for the differences and averages between marks scored in SAQ and MCQ,

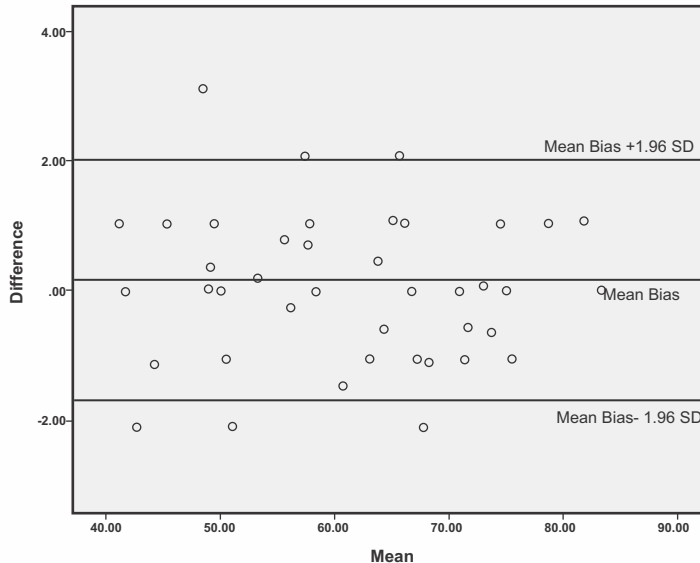


Fig.3 :Bland Altman Plot for the differences and averages between marks scored in SAQ and MCQ depicting considerable agreement between the two methods of assessment

Discussion:

Our study was designed with the hypothesis that small answer type open ended questions cannot be as objective as MCQs. In our study, we assessed the students by both open ended and closed ended questions on the same sub-topics. Contrary to our hypothesis, it was found that there was not much difference in the average scores obtained by the students, giving the impression that both the format are comparable in efficacy. Significant positive correlation was found between students' performance on MCQs and SAQs, emphasizing the fact that students who are bright, performed equally good in both the format. These findings were comparable to the findings by Dagogo J Pepple *et al* (10) who compared MCQs scoring with long essay questions. Similarly positive correlation between student performance on MCQ and short essay questions was observed by Mujeeb *et al* in Pharmacology. However, the correlation was not seen in the scores of the students who either failed or scored a distinction in the subject (11). We didn't correlate the marks in the two format on the basis of students' level of performance.

The Bland Altman plot for the measurement of agreement investigates any possible relationship between measurement error and true value. Bland Altman/Tukey mean-difference plot of the marks obtained by the two methods are showing agreement between the two methods.

A written assessment method

should be reliable, valid, cost effective and acceptable (1). There always have been a debate on type of questions most suited for a reliable assessment. Open ended questions are more pliable, requiring creativity, spontaneity—but they have lower reliability (1,3). Answering open ended questions is much more time consuming than answering multiple choice questions, hence they are less suitable for broad sampling. They are also expensive to produce and to score. Since the students have to frame the answers spontaneously, open ended questions are believed to be suitable to test the ability to solve medical problems(4,5,11). 'Subjected to examiners bias or lack of inter-examiner reliability' is probably the most recognisable disadvantage of such questions(7, 8). But in our study we have shown that if meticulously framed, open ended questions can be as objective as MCQs, free from other pitfalls of open ended questions as well.

Multiple choice questions are well known, and there is extensive experience worldwide in constructing them. Their main advantage is the high reliability per hour of testing—mainly because they are quick to answer—so a broad domain can be covered, free from examiners bias (1,3). However a common prevailing misconception about MCQs is that they are not suitable for testing the ability to solve medical problems (12). The reason behind this assumption is that all a student has to do in a multiple choice question is recognise the correct answer and the belief that MCQs test just the factual recall (13).

This was disproved by Palmer EJ and Devitt (14) who showed that the percentage of questions testing factual recall is same in MCQs as that of modified essay questions (MEQs) (14). Recently Moeen-uz-Zafar & Badr Aljarallah concluded in their study that a well-constructed MCQ is superior to MEQ in testing higher cognitive skills (15). If constructed well, multiple choice questions can test much more than simple facts as shown in the MCQs framed by us (assertion-reasoning type). This finding was supported by Hift who debated that MCQ format is better than open ended question format and suggested phasing out of open ended question format in summative assessment (16). Even if corrected for random guessing component, MCQs may overestimate some group of students and underestimate others (17). This was disproved in our study where the score in the two methods were comparable.

Research has repeatedly shown significant differences in the scores of the students with variation in the question's format (9, 16-18). However, others have proved that the question's format is of limited importance and that it is the content of the question that determines almost totally what the question tests (19,20). Every format has its advantages and disadvantages and a combination of different format is thus essential in assessing the students' performance (21). We too agree that though a well-designed testing format does not affect the performance of the students but assessment programme should include

different type of questions appropriate for the content being tested. At different stage during the course covering the subject curriculum.

Conclusion:

It is a well-known fact that MCQs are the most objective method of assessment testing higher order cognitive skills and several topics can be simultaneously covered through these questions. However, cheating and guessing component are the common drawbacks associated with this format. In this preliminary study, our hypothesis that open ended short answer type questions cannot be as objective as multiple choice type questions was proven wrong. The results suggest that meticulously-framed open-ended short answer type questions can be as objective as multiple choice questions (MCQs). Further they have short precise answers, so multiple topics can be covered and chances of guessing the answers are negligible.

Acknowledgement:

The authors acknowledge Dr. Sabyasachi Sircar, Professor & Head, Department of Physiology, for his valuable guidance in designing this study.

Conflict of interests:

None to be declared.

REFERENCES:

1. Schuwirth LWT, van der Vleuten CPM (2003). ABC of learning and teaching in medicine: Written assessment. *BMJ* **326(7390)**: 643–645.
2. Bhagat O, Bhandari B, Mehta B, Sircar S (2014). Objective Structured Practical Examination and Conventional Practical Examination: a Comparison of Scores. *Med Sci Educ* **24(4)**:395–399.
3. Schuwirth LWT, van der Vleuten CPM (2004). Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* **38(9)**:974–979.
4. Scoring rubric development: validity and reliability. Moskal, Barbara M & Jon A. Leydens [Internet]. [cited 2016 May 9]. Available from : <http://pareonline.net/getvn.asp?v=7&n=10>
5. Pinckard RN, McMahan CA, Prihoda TJ, Littlefield JH, Jones AC (2009). Short-answer examinations improve student performance in an oral and maxillofacial pathology course. *J Dent Educ* **73(8)**:950–961.
6. Rabinowitz HK, Hojat M (1989). A comparison of the modified essay question and multiple choice question formats: their relationship to clinical performance. *Fam Med* **21(5)**:364–367.
7. Anatol T, Hariharan S (2009). Reliability of the evaluation of students' answers to essay-type questions. *West Indian Med J* **58(1)**:13–16.
8. Wakeford RE, Roberts S (1979). A pilot experiment on the inter-examiner reliability of short essay questions. *Med Educ* **13(5)**:342–344.
9. Bland J, Altman D (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1(8476)**:307–310.
10. Pepple DJ, Young LE, Carroll RG (2010). A comparison of student performance in multiple-choice and long essay questions in the MBBS stage I physiology examination at the University of the West Indies (Mona Campus). *Adv Physiol Educ* **34(2)**:86–89.
11. Mujeeb AM, Pardeshi ML, Ghongane BB (2010). Comparative assessment of multiple choice questions versus short essay questions in pharmacology examinations. *Indian J Med Sci* **64(3)**:118–124.
12. Rabinowitz HK (1986). The modified essay question: effect of

- author location on student performance. *Med Educ* **20(4)**:318–320.
13. McCloskey DI, Holland RA (1976). A comparison of student performances in answering essay-type and multiple-choice questions. *Med Educ* **10(5)**:382–385.
 14. Palmer EJ, Devitt PG (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Med Educ* **7**:49.
 15. Moeen-Uz-Zafar, Badr-Aljarallah (2011). Evaluation of mini-essay questions (MEQ) and multiple choice questions (MCQ) as a tool for assessing the cognitive skills of undergraduate students at the Department of Medicine. *Int J Health Sci* **5(2 Suppl 1)**:43–44.
 16. Hift RJ (2014). Should essays and other “open-ended”-type questions retain a place in written summative assessment in clinical medicine? *BMC Med Educ* **14**:249.
 17. Forsdyke DR (1978). A comparison of short and multiple choice questions in the evaluation of students of biochemistry. *Med Educ* **12(5)**:351–356.
 18. Hettiaratchi ES (1978). A comparison of student performance in two parallel physiology tests in multiple choice and short answer forms. *Med Educ* **12(4)**:290–296.
 19. Adewoye EO, Oyebola OD, Bamgboye EA (2000). A comparative study of students' performance in preclinical physiology assessed by short and long essays. *Afr J Med Med Sci* **29(2)**:155–159.
 20. Oyebola DD, Adewoye OE, Iyaniwura JO, Alada AR, Fasanmade AA, Raji Y (2000). A comparative study of students' performance in preclinical physiology assessed by multiple choice and short essay questions. *Afr J Med Med Sci* **29(3-4)**:201–205.
 21. Anyika EN, Anyika DI (2008). Essay, multiple-choice (MCQ) and combined (essay with MCQ) type examinations: the pharmacy students' perspective. *Niger Q J Hosp Med* **18(1)**:12–15.