Original Research Article

A study on examiner variability in assessment by short essay type questions

J Prathyusha Rao[1], Cheeti Srinivas Kalyan Rao[2*], P V Chalam[3]

[1]Assistant Professor, Department of General Medicine, Bhaskar Medical college, Moinabad, INDIA.
[2]Associate Professor, Department of Paediatrics, Bhaskar Medical college, Moinabad, INDIA.
[3]Professor, Department of General Surgery, Principal, Bhaskar Medical college, Moinabad, INDIA.
Email: kalyansrinivas77@gmail.com

Abstract

Assessment of answer scripts of students is invariably associated with examiner variability. The present study is aimed to analyze the marks awarded for answer scripts of short essay type of questions for examiner variability. The answer scripts consisting of short essay type of questions of final year post graduate students in general surgery at two government medical colleges were assessed independently by two faculty members. The marks awarded were analyzed for variability for each question and total marks using 't' test and correlation for each answer between the two examiners was calculated using Spearman's correlation. The results showed no significant difference in mean scores between the examiners for most of the individual questions and total marks awarded for the answer scripts and also showed a strong monotonic correlation between the examiners but a significant difference was found in the mean scores between the examiners for two questions. Thus short essay questions are reliable and can be used to assess the cognitive skills. The current study recommends appropriate measures to decrease inter examiner variation by identifying the specific questions of variation.
Key Word: Answer scripts, Assessment, Examiner variability, Short essay questions

***Address for Correspondence:**
Dr. Cheeti Srinivas Kalyan Rao, Associate Professor, Department of Paediatrics, Bhaskar Medical college, Yenkapally, Moinabad, Ranga Reddy, Telangana – 50075.
Email: kalyansrinivas77@gmail.com
Received Date: 10/01/2019 Revised Date: 26/02/2019 Accepted Date: 12/04/2019
DOI: https://doi.org/10.26611/1004625

Access this article online

Quick Response Code:

Website: www.medpulse.in

Accessed Date: 19 April 2019

INTRODUCTION

Assessment of students in certifying examination involves assessment of different types of skills. At present short essay type of questions are widely being used to assess cognitive skills in certifying University examinations. Marks awarded should reflect the true quality and content of the answer script regardless of who and how it is assessed. Awarding marks to essays is a challenging task, as it involves subjective measures of quality and may result in variation between the assessors[1].

The marks awarded need to be reliable for meaningful interpretation. A single assessment of open ended essay type of questions has the drawback of marker subjectivity and are unreliable[1,2]The variation could be due to difference between different markers, the same marker at different times, ambiguity in question etc[1]. As reliability can be improved by multiple assessments[2], double marking of the answer scripts has been introduced at University of Health Sciences in 2011. Although double marking may address the marker subjectivity, several other factors may still influence assessment of answer scripts and awarding the mark[2].

MATERIAL AND METHODS

Institutional Ethics committee approval was obtained prior to the study. The answer scripts of exams conducted in December 2014 for final year post graduate students in general surgery at two government medical colleges, Gandhi Medical College and Osmania Medical College were included for this study. The question paper was set by the University of Health Sciences in lines of final university exams. The question paper consisted of 10

How to site this article: J Prathyusha Rao, Cheeti Srinivas Kalyan Rao, P V Chalam. A study on examiner variability in assessment by short essay type questions. *MedPulse – International Medical Journal.* February 2019; 6(2): 39-41. http://www.medpulse.in

questions of ten marks each. Thirteen answer scripts from each institution, so 26 answer scripts from both the colleges, consisting of ten answers each i.e., a total of 260 answers were included. The answer scripts were assessed independently by two faculty members who were post graduate examiners. The marks awarded were tabulated and analyzed for variability for each question and total marks. The average mark for each answer by two examiners was analyzed and mean difference in the scores of each answer was calculated using't' test. Correlation for each answer between the two examiners was calculated using Spearman's correlation. P value less than 0.05 was considered significant.

OBSERVATIONS AND RESULTS

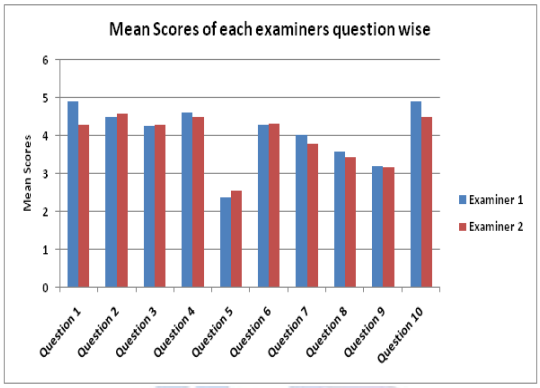

Table 1: Showing Question wise examiner mean scores, mean difference and Spearman's Correlation

	Mean Score Examiner 1	Standard Deviation Examiner 1	Mean Score Examiner 2	Standard Deviation Examiner 2	Mean difference of scores (t-test) P value	Spearman's Rho (Examiner 1 & 2)	Spearman's Rho Significance (P value)
Question 1	4.904	1.5364	4.308	1.6191	.014	0.716	<0.001
Question 2	4.500	1.7321	4.583	1.5012	.790	0.56	0.004
Question 3	4.261	1.4684	4.304	1.4596	.829	0.692	<0.001
Question 4	4.604	1.2682	4.500	1.4744	.585	0.757	<0.001
Question 5	2.381	1.3314	2.571	1.6605	.515	0.559	0.008
Question 6	4.292	1.1602	4.333	1.2039	.824	0.684	<0.001
Question 7	4.038	1.3995	3.808	1.2335	.228	0.727	<0.001
Question 8	3.580	1.6813	3.440	1.3254	.488	0.774	<0.001
Question 9	3.196	1.5575	3.174	1.4350	.937	0.63	0.001
Question 10	4.917	1.2306	4.500	.7802	.038	0.576	0.003
Total for all questions	38.038	11.7932	37.000	10.3383	.403	0.755	<0.001

The analyzed data of Question wise examiner mean scores, mean difference and Spearman's Correlation are shown in the Table 1. The results show that there is no significant difference in mean scores between the examiners for questions 2 to 9. There is significant difference in mean scores between the examiners for questions 1 and 10. There is no significant difference in mean scores between the examiners for total marks awarded for the answer script. The Spearman's Rank Correlation show a positive monotonic correlation between the examiners for all the questions with a statistically significant p value of <0.05. Questions 1, 4, 7 and 8 had Spearman's correlation of more than 0.70 indicating a stronger monotonic correlation. There was strong monotonic correlation between the examiners for the total marks awarded with Spearman's correlation of more than 0.75.

DISCUSSION

In this study there is significant difference in mean scores between the examiners for two questions out of ten with strong monotonic correlation between the examiners for the total marks awarded for the answer script. The variation in the scores between the examiners for the answers could be due to multiple factors. A qualitative study has shown that markers have internal concerns about their ability to mark fairly and dealing with pass/fail borderline scripts and the consequences of the mark on the student were particular concerns[3]. Meticulously framed open ended short answer type

questions can be as objective as multiple choice questions[4].It is essential that each of the question need to be structured and validated to avoid variation in the assessment. Some of the methods suggested to decrease the variation are double marking [5], introduction of guidelines to the examiners and preparation of model answer paper[6]. Thus short essay questions can be more reliable by decreasing the variation if the cause for variation between examiners in awarding the marks to specific question answers can identified and bring in appropriate modifications.

CONCLUSIONS

There is no significant variation in the total marks awarded to an answer script of short essay type of questions. Thus short essay questions are reliable and can be used to assess cognitive skills. There is significant variation between examiners in awarding marks for two out of ten question answers. Further research is needed to identify the cause for variation between examiners in awarding the marks to specific question answers and bring in appropriate modifications to decrease the variation.

ACKNOWLEDGEMENT

Pavithra Ramamurthy Umesh, Medical student, Oregon Health and Science University for helping in the literature search.

Dr. Massarat Begum, Assistant Professor, Physiology, Bhaskar Medical college, Moinabad for the preparation of final manuscript.

REFERENCES

1. T. Anatol, S. Hariharan. Reliability of the Evaluation of Students' Answers to Essay-type Questions. West Indian Med J 2009; 58 (1):13-16.
2. R. E.Wakeford, Sarah Roberts. Short answer questions in an under gradute qualifying examination: a study of examiner variability. Medical Education 1984; 18: 168-173.
3. Kamila Hawthorne, Fiona Wood†, Kerenza Hood, Rebecca Cannings-John and Helen Houston. Learning to mark: a qualitative study of the experiences and concerns of medical markers. BMC Medical Education 2006, 6: 25.
4. Bharati Mehta1, Bharti Bhandari1, Parul Sharma2, Rimplejeet Kaur3 Short Answer Open-Ended versus Multiple-Choice Questions: A Comparison of Objectivity. Ann Natl Acad Med Sci (India), 52(3): 173-182, 2016
5. Cannings R, Hawthorne K, Hood K, Houston H. Putting double marking to the test: a framework to assess if it is worth the trouble.Med Educ. 2005 Mar; 39(3):299-308.
6. R.E.Wakeford, Sarah Roberts. A pilot experiment on the inter-examiner reliability of short essay questions. Medical Education,1979, 13, 342-344.

Source of Support: None Declared
Conflict of Interest: None Declared