

Discriminant Validity and Reliability of Scores of Multiple Choice and Short Essay Questions

Anila Jaleel¹, Zohra Khanum², Imran A Siddiqui³, Mohsin Ali⁴, Saba Khalid⁵, Rakhshan Khursheed⁶

ABSTRACT:

Background and Objective: Assessment of learning plays a vital role in curriculum allowing faculty to evaluate the achievement of student and educational process. Quality assurance of the assessments is done to assess the quality of teaching and learning. A study is therefore conducted to find the discriminant validity and reliability of scores of Multiple Choices and Short Answer Questions.

Methods: A correlation study was conducted on 272 second year MBBS students of Fatima Memorial College of Medicine and Dentistry, Lahore from 2016 – 2017. Students were assessed in end of year examination with multiple choice questions (MCQs) and short essay questions and (SEQs) in the subjects of Physiology and Biochemistry. A total of 149 second-year MBBS students took the Physiology and 143 students took the end of year Biochemistry examination with 50% minimum passing level. The mean and standard deviation of the scores were calculated and the scores of MCQs and SEQs were correlated by applying Pearson's correlation. Reliability was determined by Cronbach's alpha. Discriminant validity of scores (MCQs and SEQs) was analyzed by Pearson's correlation.

Results: Students scored significantly higher in MCQs than SEQs in end of year examination of Biochemistry and Physiology. Positive significant correlation of scores (MCQs and SEQs) was obtained in Biochemistry while correlation was positive but non-significant in Physiology. The reliability assessed by Cronbach's alpha was moderate in Biochemistry and Physiology assessments. Discriminant validity was determined in both subjects by determining the correlation between the scores of the MCQs and SEQs, constructed according to the same module.

Conclusion: The internal assessment of students based on end of year examinations in the subjects of Physiology and Biochemistry scored moderate in terms of reliability and discriminant validity was found between scores of MCQs and SEQs in each module.

KEYWORDS: Discriminant Validity, Assessment, Reliability, Multiple choice questions, Short essay questions.

How to Cite This:

Jaleel A, Khanum Z, Siddiqui IA, Ali M, Khalid S, Khursheed R. Discriminant validity and reliability of scores of multiple choice and short essay questions. *Biomedica*.2020; 36 (2):193-8.

-
1. Prof. Dr. Anila Jaleel
Professor of Biochemistry Department
Shalamar Medical and Dental College
University of Health Sciences, Lahore – Pakistan.
 2. Prof. Dr. Zohra Khanum
MBBS, FCPS, DCPS, MCPS (Medical education)
Professor, Obstetrics and Gynae department
Fatima Jinnah Medical University, Lahore – Pakistan.
 3. Dr. Imran Siddiqi
MBBS, M. Phil, Ph.D, MHPE
Associate Professor
Department of Medical Education TCU and UNTHS
School of Medicine, Fort Worth, USA.
 4. Dr. Mohsin Ali
MBBS, M. Phil
Associate Professor of Physiology
Abu Ammara Medical College, Lahore- Pakistan.
 5. Prof. Dr. Saba Khalid
Professor, Department of Physiology
Fatima Memorial College of Medicine and Dentistry, Lahore – Pakistan.
 6. Dr. Rukshan Khursheed
Assistant Professor, Biochemistry Department
Shalamar Medical and Dental College, Lahore – Pakistan.

Corresponding Author:
Prof. Anila Jaleel
Head & Professor Department of Biochemistry
Shalamar Medical and Dental College, Lahore – Pakistan.
Email: aneelajaleel@hotmail.com

- Received for publication: 16-02-2020
- First revision received: 19-04-2020
- Second revision received: 03-06-2020
- Accepted for publication: 15-06-2020

INTRODUCTION

Both assessment for learning and assessment of learning play an important role in learning. Several studies have shown item analysis as part of quality assurance in assessments with determination of difficulty and discriminatory index in MCQs, relationship of item difficulty with nonfunctional distractors and have developed faculty development programs to improve the quality of assessments.¹⁻⁴ An important aspect of ensuring quality assurance for an examination is to identify the relationship amongst the variables—that is, the correlation between one assessment tool with the outcome of the other tool measuring the same skill. Construct validity of scores is able to capture the latent variable.⁵ To maintain the high quality of medical education; students are assessed using multiple assessment tools. Such a high stakes examination necessitates ensuring the construct validity of the tools. Construct validity includes convergent and discriminant validity. Convergent validity determines whether constructs that are related to each other are actually related; namely, theoretically similar constructs should be correlated.⁶ However, discriminant validity tests unrelated constructs for which no relationship is expected to ensure that there is in fact no relationship.⁷ It points out that the items of one construct type should correlate higher amongst themselves than with the others, which are theoretically not supposed to correlate.⁵

Study by Mahmood (2015) shows moderate correlation between the MCQs and SEQs, however inter-rater reliability of SEQs showed high correlation.⁸ Similar study was done by Farooqui et al.⁹ to assess the correlation of MCQs and SEQs in medicine, surgery, pediatrics and obstetrics and gynecology examinations. Moderate positive significant correlation was found between MCQs and SEQs in three subjects except for obstetrics and

gynecology where weak positive correlation was found.

Multiple tools for assessment are employed at end of year examinations in medical colleges affiliated with University of Health Sciences as per the guidelines. Medical Education department is being established in every medical college to ensure quality assurance in medical education. This study was conducted to find out the discriminant validity and reliability of tools of assessment in written examination i.e. MCQs and SEQs and correlation between them in end of year examination in private medical college affiliated with UHS.

METHODS

A correlation study was conducted to determine the discriminant validity of MCQs and SEQs scores in end of year examination on 272 second year MBBS students, Fatima Memorial College of Medicine and Dentistry, Lahore from 2016-2017. Students were assessed with MCQs and SEQs in the subjects of Physiology and biochemistry. Objective structured practical exams (OSPE), viva voce exams and practical exams were excluded from the study. Second-year MBBS students took the Physiology (149) and Biochemistry examinations (143) with 50% minimum passing level as absolute standard setting. Discriminant validity was calculated on scores of MCQs and SEQs in a specific module given in end of year examination according to table of specification so that content of the two match each other. Scores of MCQs and SEQ in all modules in second year MBBS were analyzed.

The examination consisted of 45 MCQs with a maximum possible score of 45 marks, based on the test blueprint provided by the University of Health Sciences Lahore, and nine SEQs worth a total of 45 marks. The duration of the two exams was 2.5 hours (one hour for the MCQs and one and a half hours for the SEQs) for each discipline with minimum passing level 50%. Ethical approval for this study was granted by the Institutional Review Board of FMH College of Medicine and Dentistry vide letter No: FMH-05-IRB-254-M dated May 19, 2017 and Ethical review committee University of Health Sciences dated 15th April 2016, prior to data collection.

STATISTICAL ANALYSIS

Data was analyzed by SPSS 20.0. The mean and standard deviation of the scores was calculated and the scores of MCQs and SEQs were correlated by applying Pearson's correlation. Chi-square test was applied to determine the significance of the students passed or failed in MCQs versus SEQs in both subjects. Reliability was determined by Cronbach's alpha. P value <0.05 was taken as significant.

RESULTS

Performance of Students in SEQs and MCQs in Biochemistry and Physiology based on passed or failed the end of year written examination is shown (Fig.1). Out of 149 students who took Physiology examination, 89.9% were declared passed. Significantly high number of students passed both the SEQ and MCQ. Similarly, in Biochemistry 113 students out of 143 were declared passed. Both the MCQ and SEQ were passed by 69 students while 23 failed both the MCQ and SEQ. The Pearson's Chi-square test was applied to determine the significance of the results. Significantly more students passed the MCQ than the SEQ in the subjects of Physiology and Biochemistry ($P = 0.05$ and $P = 0.001$ respectively) (Fig.1).

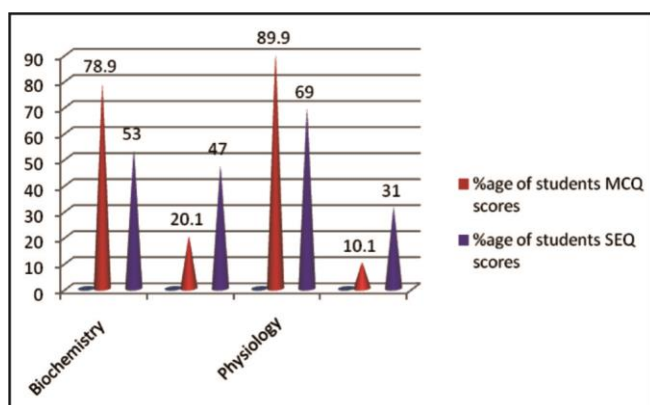


Fig.1: Performance of Students in SEQs and MCQs in Biochemistry and Physiology.

Mean scores of MCQs and SEQs in the subjects of Biochemistry and Physiology are tabulated as Table-1. Mean scores of MCQs was higher than SEQs in both subjects. Scores of MCQ and SEQ correlated significantly in both subjects (Table-1).

Table-1: Mean and standard deviation with correlation of scores of MCQs and SEQs in Biochemistry and Physiology.

	MCQ Scores*	SEQ Scores*	Correlation of Scores (r value)	P value
Biochemistry** (n=143)	27.22 ± 5.62	22.65 ± 6.41	0.59	0.047
Physiology** (n=149)	28.60 ± 4.91	24.64 ± 5.38	0.39	0.049

*Values expressed as mean + SD

** No of students

Reliability or internal consistency of end of year examination in Biochemistry and Physiology is tabulated as Table-2. The Biochemistry examination consisted of 45 MCQs and 9 SEQs (18 items). Cronbach's alpha was applied, which showed acceptable reliability i.e. 0.72. The Physiology examination consisted of 45 MCQs and 9 SEQs (17 items) and reliability was found to be low i.e. 0.61. The minimum acceptable value for Cronbach's alpha is 0.70. Less than 0.70 is taken as low.

Table-2: Reliability (Internal Consistency) of Mock Examination in Biochemistry and Physiology.

Subjects	Mock Examination	Cronbach's Alpha Value
Biochemistry	MCQs (45 items)	0.33
	9 SEQs (18 items)	0.91
	Both MCQs and SEQs	0.72
Physiology	MCQs (45 items)	0.50
	SEQs (17 items)	0.85
	Both MCQs and SEQs	0.61

Discriminant validity of the scores for the MCQs and SEQs in Biochemistry and Physiology, respectively are tabulated as table 3 and 4. Discriminant validity of scores of MCQs and SEQs of specific module was analyzed to ensure the validity of similar content. Second year MBBS course of Biochemistry consists of ten modules which were analyzed individually and r value of each module value was determined. The correlation r value was non-significant in almost all modules of both subjects which shows that the constructs of MCQs differ from SEQs (Tables-3,4).

Table-3: Discriminant validity of scores of MCQ's and SEQ's Biochemistry according to Modules.

Modules	MCQs No	SEQs No	r value	P value
CHO Met	1 - 5	Q1	-0.126	0.125
Bioenergetics	25, 26	Q9(b)	-0.089	0.288
Protein Met	6 - 12	Q3	0.105	0.211
		Q7a	0.010	0.909
Lipid Met	13 - 18	Q2	0.156	0.063
		Q9(a)	0.348	0.000
Mol Biology	31 - 32, 38 - 45	Q5	0.016	0.852
GIT	33, 34	Q6a	-0.114	0.176
Oncogenesis/ Xenobiotics	29, 30	Q6b	0.023	0.788
Endocrinology	19 - 24	Q7b	0.070	0.409
Nucleic acid	35 - 37	Q4	0.001	0.994
Acid base balance	27, 28	Q8	0.040	0.632

CHO: Carbohydrate, Met: Metabolism, MOL: Molecular, GIT: Gastrointestinal tract

Table-4: Discriminant Validity of Scores MCQs and SEQs Physiology according to Modules.

Objectives	MCQs No	SEQs No	r value	P value
CNS	17 - 2	Q1	0.066	0.424
CNS	23 - 27, 33 - 36	Q7	0.006	0.940
Kidney	28, 29	Q2	-0.039	0.633
Kidney	30, 31	Q9	0.007	0.935
Reproduction	9,10,13,16,30,32	Q4	-0.024	0.775
GIT	41 - 45	Q6	-0.211	0.010
Special Senses	1 - 5, 6-8, 40	Q8	0.100	0.226
Endocrinology	37 - 39	Q5	-0.010	0.899
Endocrinology	11 - 15, 29, 35	Q3	0.156	0.058

CNS: Central Nervous system

DISCUSSION

The study evaluated the performance of students using MCQs and SEQs in basic science subjects, Biochemistry and Physiology. The reliability and validity of the scores were determined. Discriminant construct validity was found in the scores of MCQs and SEQs in both subjects. Students performed better on the MCQs compared to the SEQs. Scores on MCQs and SEQs showed positive significant correlation in both subjects. The internal consistency of SEQs in Biochemistry and Physiology was adequate, but low in the MCQ examination for both subjects. Moderate consistency was found for the combined written examinations in Biochemistry but was low in Physiology.

According to present study, the internal consistency in scores of MCQs for both basic subjects was low. The reasons could be that there

were fewer items on the specific objective being tested, or examination items were either too difficult or too easy. It is suggested that the use of up to 75% of items of average or moderate difficulty may help to improve reliability. Moreover, using items in other tests and then correcting or deleting them from the bank after post-examination psychometric analysis, will improve the reliability of future examinations.¹⁰ Present data offer a need to focus on and improve the reliability of the assessments by post-examination analysis and the test-retest strategy.

The internal consistency in scores of SEQs for both subjects in current study was rated adequate. It showed that following the specifications, provision of structured keys and scores by two different examiners for inter-rater reliability improved the reliability of the SEQ exam in both subjects.

Students performed better on the MCQs than on the SEQs, and scores of both tools correlated significantly in current study, while those students who scored better on the MCQs also scored well on the SEQs. However, the two methods are different yet interrelated in terms of assessing the cognitive level of the students. This shows that students with conceptual knowledge of the subject can perform well regardless of the assessment tool. Teaching modalities should emphasize deep learning strategies and building concepts rather than rote learning. However, a study found mild to moderate correlation between scores of SEQ's and MCQs for students of final-year M.B.B.S. Author suggested that additional studies are needed to determine the correlation between these two modes of examination and increase the validity of assessments via SEQ's.⁹ Moderate positive correlation ($r = 0.5$ $P < 0.01$) between scores of MCQs and SEQs was found in the send up undergraduate examination for Ophthalmology.⁸

On the other hand a study showed significant positive correlation of performance between SEQs with MCQs in the subject of Pharmacology at Kasturba Medical College at Manipal University, Karnataka, India.¹¹ Additionally a group of workers showed significant positive correlation between SEQs and MCQs for an Otolaryngology examination and concluded that performance was independent of the question type.¹² Another study also reported correlation between MCQs and SEQs, indicating

that students performed well in both MCQs and SEQs.¹³

According to present study non-significant correlation in scores of MCQs and SEQs was found in specific modules (based on test blueprints). The correlation shows discriminant validity, which revealed that SEQs and MCQs have different constructs across learning domains and assess the varied depth of the content. It should not be considered that both serve the same purpose. Present study is in agreement with a study that suggested that quality of assessment tools should be improved, as the measurement of assessment tools of low cognitive rank not only produce reduction in the validity of the test but also force the medical students to use surface or unreflective learning: unsuitable for self-motivated learning.¹⁴ Moreover, as the construct of each is different, students are assessed on different aspects of written examinations. The MCQs are structured as answer keys are available and the questions themselves provide the information. While SEQs, which are also structured, require students to interpret and analyze to provide the answer as well as evaluate students on the basis of written communication skills. Study proved that these two constructs are different and assess different learning domains.

CONCLUSION

The study showed that students performed better on MCQs than they did on SEQs. Positive significant correlation was found in scores of MCQs and SEQs but discriminant validity was found between them when analyzed separately on each module. This showed that students who performed well on MCQs also scored higher on SEQs; while the assessed contents are same though their constructs are entirely different and are used to assess varied depth of knowledge.

LIMITATION OF STUDY

The sample size of the study was small as only second-year MBBS students were included. Inclusion of MBBS students across four years and more basic sciences subjects would reveal more detailed facts.

ACKNOWLEDGEMENT

The authors owe thanks to Mr. Khalid Biostatistician (FMH college of Medicine and Dentistry) for statistical guidance in the study.

CONFLICT OF INTEREST

None to declare.

FINANCIAL DISCLOSURE

None to disclose.

REFERENCES

1. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology. *South-East Asian J Med Educ.* 2013; 7 (1): 45-50.
2. Chavda P, Misra S, Duttaroy B. Item analysis of multiple-choice questions based undergraduate assessment in community medicine. *South East Asian J Med Educ.* 2015; 9 (1): 66-8.
3. Abdulghani HM, Ahmad F, Ponnampereuma GG, Khalil MS, Aidrees A. The relationship between non-function distracters and item difficulty of multiple-choice questions: A descriptive analysis. *J Health Spec.* 2014; 2 (7): 148-51.
4. Abdulghani H M, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK, Syed S, et al. Faculty development programs improve the quality of multiple-choice questions items' writing. *Sci Rep.* 2015; 5 (5): 9556-61.
5. Zait A and Bertia PE. Methods for testing discriminant validity. *Manage & Market J.* 2011; 9 (12): 217-24.
6. Delaram M. and Sharifi A. A comparison of students, scores in multiple-choice and essay questions in mother and child health examination. *Future Med Educ J.* 2014; 4 (2): 15-8.
7. Hassan S, Matpa MN and Yusoff MSB. Discriminant and convergent validity of measurement tools in postgraduate medical education of a surgical-based discipline: Towards assessment program. *Edu Med J.* 2012; 4 (1): e22-42.
8. Mahmood H. Correlation of MCQ and SEQ scores in written undergraduate ophthalmology assessment. *J Coll Phys Surg Pak.* 2015; 25 (3): 185-8.

9. 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.
10. Pai M, Sanji N, Pai PG and Kotian S Comparative assessment in Pharmacology multiple choice questions versus essay with focus on gender differences. *J Clin Diag Res*. 2010; 4 (7): 2515-20.
11. Mujeeb AM, Pardeshi ML, Ghongane BB. Comparative assessment of multiple-choice questions versus short essay questions in Pharmacology examinations. *Indian J Med Sci*. 2010; 64 (3): 118-24.
12. Sharma H and Mutalik M. Comparison between student scores in multiple choice questions and structured essay questions at III MBBS examination in otorhinolaryngology in a medical college in Mauritius. *Int J Pham Bio Sci*. 2014; 5 (4): 1195-1202.
13. Baig M, Ali SK, Ali S, Huda N. Evaluation of Multiple Choice and Short Essay Question items in Basic Medical Sciences. *Pak J Med Sci*. 2014; 30 (1): 3-6.
14. DiBattista D and Kurzawa L. Examination of the quality of multiple-choice items on classroom tests. *CJSoTL*. 2011; 2 (2): 1-4.

Author's Contribution

AJ: Conception and design of study, acquisition of data.

ZK: Conception and design of study.

IAS: Data analysis and interpretation.

MA, SK: Acquisition and compilation of data of Physiology.

RK: Drafting the manuscript with intellectual input.

ALL AUHTORS: Final approval of the version to be published.