

Test Item Analysis of MCQS of Medical Physiology: Summative Assessment

Abdul Rehman Khokhar¹, Qurrat-ul Ain Rehman², Mujahid Hussain³

¹Department of Physiology, Ghazi Khan Medical College, Dera Ghazi Khan, Pakistan

²Department of Biology, PAEC Inter Science College, Dera Ghazi Khan, Pakistan

³Department of Statistic, Government Associate College, Dera Ghazi Khan, Pakistan

Email address:

Prof.arkhokhar@gmail.com (A. R. Khokhar), arehmankhokhar777@gmail.com (A. R. Khokhar)

To cite this article:

Abdul Rehman Khokhar, Qurrat-ul Ain Rehman, Mujahid Hussain. Test Item Analysis of MCQS of Medical Physiology: Summative Assessment. *Biomedical Sciences*. Vol. 8, No. 1, 2022, pp. 37-40. doi: 10.11648/j.bs.20220801.16

Received: February 4, 2022; **Accepted:** February 24, 2022; **Published:** March 3, 2022

Abstract: *Introduction:* MCQs in one of popular test item of educational institution in a short time, large part of curriculum. It is an effective tool for student assessment as well as providing guidelines to teachers. *Objectives:* Our study objectives were to analyze the quality of MCQs and identify the low scorers and student learning difficulties. Standardization of MCQs is required only after test item analysis. *Material and Methods:* The study design was cross sectional analytical study. Simple random sampling technique was applied. The total participants were 100 students of 1st year MBBS of Ghazi Khan Medical College, Dera Ghazi Khan. Fifty best MCQs were given in send up of Medical Physiology 2021. On basis of students total score in test, they were divides into two groups, high scorers (n=25) and low scorers (n=25). The difficulty index (P) and discrimination index, distracter efficiency (DE%) were calculated. *Results:* Our study 80% of test items have acceptable range of difficulty index and very high distracter efficiency. Only 20% of test item showed very poor difficulty index, which needs to be rephrased or deleted. The 20% test item had acceptable difficulty index (P=38), high D. I (0.36) and highest distracter efficiency (DE=100%). *Conclusion:* Test items with average difficulty index (P), excellent DI and all functional distracters should be given in subsequent tests.

Keywords: Test Item, Difficulty Index, Discriminative Index, Distracter Efficiency

1. Introduction

One best answer (A-Type) questions are the most widely used multiple choice items. MCQs consist of an item, followed by a series of choices, having one correct answer and four distracters. The test item flaws like absolute terms, long correct answers, grammatical and, logical cues, word repeats, convergence strategy must be avoided. Irrelevant difficulty can be eliminated by avoiding long/complicated, options and in chronological order, vague terms e.g. rarely, usually, numerical data, options like none of the above, tricky items, hinged answers [1]. The quality of test item can be analyzed by classical test theory and test response theory (IRT) which has filled the gaps and ideal for test item analysis [2].

MCQs in one of popular test item of Professional education, in a short time, large part of curriculum can be

assessed and less effort on student part, although a high quality one best MCQs need much effort and time by the examiner. It is an effective tool for student assessment as well as providing guidelines to teachers [3]. Designing one best MCQ is a challenging task and time consuming for examiners. Only after testing MCQs can be standardized. Test item analysis consists of student responses to individual test items and test as a whole [4].

More effort is needed to writing best MCQs than writing a good question. A properly constructed MCQs can assess higher cognitive process of Bloom's taxonomy like interpretation, synthesis, and practical application of knowledge, rather than just testing recall of isolated and essential facts [5]. Item analysis evaluate the assessment tool which is beneficial both for student and teacher. The

reliability and validity of test item is determined by post examination analysis of student responses [6]. MCQs are used strategically to test important content and mixed with the practical testing of clinical competence. MCQs may result testing fair and valid. It is true that no single format should be used for assessment exclusively. American Educational Research Association and National Council of Education recommended that MCQs testing is a commonly used format for both formative and summative assessments [7]. Subali *et al.* [4] reported that test item difficulty index and easy items are easily identified. It is need of time to identify test item characteristics and associated factors for development of good test item [8].

Test item analysis provides feedback to teachers to install changes in teaching standards. Item analysis helps to find out poor item which need improvement or deletion. What was the quality of our MCQs? Test item analysis is a reliable test to discriminate the student performance in examination. Researchers reported that it is not a practice that data generated have not still utilized to develop subsequent MCQs item [9, 10]. Difficulty index (P-Value) is percentage of students who answered the item correctly, which is a 0-100%. The recommended difficulty index is 30-70%. Easy items need be placed either at start of test or deleted. While difficult items should be reviewed for confusing language, or even incorrect key [11]. Discriminative index (DI) is ability of item to distinguish between high and low scorers. It ranges 1.00 to + 1.00. It is presumed that high scorers select correct answer for each item more often than low scorers. DI values between 0.20-0.35 are considered good. Its value > 0.35 is taken as excellent and its value < 0.20 is considered as poor. An item contains a stem and five options including one correct key and four distracters alternatives [12].

An option, other than key, was selected by less than 5% or more students on basis of number of NFD. Effective distracter is the option selected by 5% or more students. On basis of number of NFD, efficiency of distracter to evaluated [13]. Our study objectives were to analyze the quality of MCQs, to improve the items that need correction or deletion. The study purpose was also to identify the low scores and student learning difficulties, which can be improved by counseling or by modifying teaching methods. Teachers can get feedback on their teaching efficiency, which may lead to improvement in teaching skills in the future.

2. Materials and Methods

Study was conducted at Ghazi Khan Medical College, Dera Ghazi Khan during September, 2021. It was a cross sectional analytical study. Total participants were 100 students of MBBS 1st year of Ghazi Khan Medical College Dera Ghazi Khan. Fifty best MCQs were given in Send Up examination of physiology, topics included were cell, nerve and muscle, blood, respiration, cardiovascular system, body temperature regulation and human body adjustment to different environments. MCQs items were constructed and vetted by Head of Physiology department. The time given for MCQs paper was 50 minutes and followed by SEQs paper. The MCQs having a stem and five options, one of them was correct and other four were distracters. Students had to choose the correct answer and mark on given MCQs answer sheet. Each correct response was given one mark. No mark was awarded for incorrect or blank response. Maximum score of overall test was 100 and minimum was 0, and no negative marks were allotted. After randomization test item were selected by lottery method, were analyzed and, optimized before adding to MCQs bank.

Steps for item analysis were,

- Scoring whole test for all students
- Rank students on basis of their test scores
- Top ¼ were taken as high achievers (h) and bottom ¼ were taken as low scorers (l)
- Tables were prepared for each item and calculation made to determine, difficulty index, discriminative index and distracters efficiency by using formulas [14].

$$\text{Difficulty Index (P)} = h + l / n \times 100$$

$$\text{Discriminative Index (DI)} = h - l / n \times 2$$

h=number of students answering correctly in high group

L=number of students answering correctly in low group

n=total number of students in two groups including non-responders.

The relationship between difficulty index and discriminative index was determined. The distracter analysis was done to check the efficiency of all distracters of test item.

Table 1. Interpretation of different values of difficulty index, discriminative index (DI) and distractor efficiency (DE) [14].

Test Item	Student Group	A Option	B Option	C Option	D Option	E Option	P	DI	DE%
6	High	1 (4%)	8 (32%)	12 (48%)	3 (12%)	0 (0%)	52	0.08	75
	Low	-12%	4 (16%)	14 (56%)	2 (8%)	0 (0%)			
8	High	1 (4%)	2 (8%)	5 (20%)	2 (8%)	14 (56%)	38	0.36	100
	Low	5 (20%)	5 (20%)	6 (24%)	3 (12%)	5 (20%)			
22	High	2 (8%)	14 (56%)	5 (20%)	1 (4%)	3 (12%)	46	0.2	100
	Low	0 (0%)	9 (45%)	3 (12%)	5 (20%)	7 (28%)			
34	High	8 (32%)	3 (12%)	4 (16%)	10 (40%)	0 (0%)	38	0.06	75
	Low	10 (40%)	4 (16%)	3 (12%)	7 (28%)	1 (4%)			
46	High	0 (0%)	9 (36%)	0 (0%)	5 (20%)	11 (44%)	15	0.08	75
	Low	0 (0%)	6 (24%)	5 (20%)	10 (40%)	3 (12%)			

Table 2. Test Item Analysis showing p value, DI and DE%.

Test Item	Student Group	A Option	B Option	C Option	D Option	E Option	P	DI	DE%
6	High	1 (4%)	8 (32%)	12 (48%)	3 (12%)	0 (0%)	52	0.08	75
	Low	-12%	4 (16%)	14 (56%)	2 (8%)	0 (0%)			
8	High	1 (4%)	2 (8%)	5 (20%)	2 (8%)	14 (56%)	38	0.36	100
	Low	5 (20%)	5 (20%)	6 (24%)	3 (12%)	5 (20%)			
22	High	2 (8%)	14 (56%)	5 (20%)	1 (4%)	3 (12%)	46	0.2	100
	Low	0 (0%)	9 (45%)	3 (12%)	5 (20%)	7 (28%)			
34	High	8 (32%)	3 (12%)	4 (16%)	10 (40%)	0 (0%)	38	0.06	75
	Low	10 (40%)	4 (16%)	3 (12%)	7 (28%)	1 (4%)			
46	High	0 (0%)	9 (36%)	0 (0%)	5 (20%)	11 (44%)	15	0.08	75
	Low	0 (0%)	6 (24%)	5 (20%)	10 (40%)	3 (12%)			

3. Results

Table 1 shows interpretation of different values of difficulty and, discriminative indices, distracters efficiency. Table 2 shows the comparison of different test items on basis of their difficulty index (P), discriminative index (DI) and distracter efficiency (DE%).

Test item no #6 having five options (A-E) was analyzed among high scorers group (n=25) and low scorers group (n=25). Its calculated difficulty index was 52, which is interpreted as acceptable. Discriminative index of this item come to be 0.08, which can be interpreted as poor. Interestingly this item shows very high distracter efficiency (DE%=75). Test item no 8 was analyzed among high scorers (n=25) group and low scorer group (n=25) of the study. Its difficulty index was 38, which can be interpreted as acceptable, while discriminative index as calculated as 0.36, which was interpreted as excellent. This test item showed highest distracter efficiency (DE%=100). Test item no # 22 was analyzed among high and low scorer groups. The difficulty index of this test item was calculated as 46, which is interpreted as acceptable. Discriminative index of this test item come to be 0.20, which was marginal. This item had highest distracter efficiency (DE%=100).

Test item No # 34 was analyzed among the high scorers and low scorer's groups. The difficulty index was calculated as 38, which is interpreted as acceptable. DI of this test item was come to be 0.06, which was poor, while it has very high distracter efficiency (DE%=75). Test item no # 46 was also analyzed among the high scores and low scorers of this study. The difficulty index was calculated as 15, it means, it was too difficult item. The DI of this item was 0.08, which can be interpreted as poor. While distracter efficiency of this test item was very high (DE%=75).

So 80% of test items showed p value of acceptable level and 20% test item were difficult.

4. Discussion

MCQs are useful assessment tools to measure factual recall, if constructed carefully can also assess high order of thinking and skills which are important for medical education. Method of assessment need regular evaluation. The validity of MCQs needs evaluation of test item, to check

how effective for assessing the knowledge of medical students. Test items with poor discriminative index (DI) and difficulty index (P), having poor distracter efficiency (DE%) should be reviewed for correction and reconstruction or deletion. Item analysis should be carried out regularly to test quality of MCQs for subsequent tests [15]. Examining body should have regular practice of evaluation of test item prior to subsequent use.

Pande et al (2013) reported that test items have maximal discrimination with difficulty index 40-60. While very easy and too difficult items had poor discrimination among high and low scores [16]. Forty percent of test item were having difficulty index between 40-60%. But 80% of test items were within acceptable range i.e. (30-70%). Current study shown difficulty index of 80% (n=4) items was within acceptable range, and 20% of test items were interpreted as poor and marginal respectively. While distracter efficiency was 100% among while 25% (n=1) item was too difficult. The DI of 20% items was excellent and 60% (n=3) 40% of test items and 40% of test item showed very high DE% (75). The 80% of study test items were within acceptable range of difficulty index, very high distracter efficiency, only 20% of study test item showed very poor difficulty index which needs to be rephrased or deleted.

MCI (2019) and Jonathon et al (2018) reported that DI >0.2 is acceptable and able to discriminate between good and weak students. Contemporary study shown 40% of test item analyzed had DI > 0.2 which was similar to above mentioned workers and no test item showed negative discrimination [17, 18].

Mitra et al (2009) reported that DI has poor correlation with difficulty index [19]. Current study findings of DI and difficulty index are in accordance with Mitra et al, 2019. Distracters are designed to evaluate student performance. The distracters efficiency (DE) rectifies the errors in distracters so that they may be revise, replaced or removed [19]. Current study 40% (n=2) test items had DE of 100% and 60% (n=3) test item had DE of 75%, that shows the effectiveness of distracters among test items. Mehta & Makhasi (2014) reported that test item having one or two NFDs should be considered better than having no NFDs [20].

Rao et al (2016), Raj Kumar et al (2018) and Bhat et al (2021) reported NFD (5%), 43.3% and 22.7% respectively which are in accordance with our study [21-23]. The test items of current study having 0-1 NFDs were shown to be good test item, having

the difficulty index (30-70%), and test item no.8 had acceptable difficulty index ($P=38$), high DI (0.36) and the highest distracter efficiency ($DE=100\%$). So among five test item studied, the test item no.8 was the best one.

5. Conclusion

Test items with average difficulty index, excellent DI and all functional distracters were recommended for Professional examination. Test item analysis should be practiced for standardization of MCQs paper, to improve teaching and learning experiences.

Abbreviations

MCI ---- Medical Council of India
 P value –Difficulty index
 DI ----- discriminative index
 DE -----Distractor efficiency
 NFD----- Non Functional Distractor

Authors Contribution

A. R. Khokhar: Conception, Data Collection, Manuscript writing

Q. A. Rehman: Study Design, Literature Review, Data interpretation

M. Hussain: Data analysis, Sampling technique, Critical review

References

- [1] Ahila E, Velou MS, Refine the multiple choice questions tool with item analysis, 2021; 7: 80-85.
- [2] Hingorjo MR, Jaleel F, Analysis of one best MCQs: The difficulty index, discrimination index and distractor efficiency. JPMA, 2012; 62: 142-7.
- [3] Juniati E, Subali B, Teacher's opinion about learning continuum of genetic based on student's level of competence AIP Conference Proceeding 1868, 100002; 2017, <http://doi.org/10.1063/1.4995212>.
- [4] Subali B, Kumaidi, Aminah NS, The comparison of item test characteristics viewed from classic and Modern Test Theory IJI 2021; 14: 647-660.
- [5] Singh T, Gupta P, Singh D, Test and item analysis. In: principles of Medical Education 3rd Ed. New Delhi Jaypee Brothers, 2009: 70-7.
- [6] Carneson J, Delpierre G, Masters K, Designing and Managing MCQs. Appendix C: MCQs and Bloom Taxonomy (online) 2021 (cited Jun 25) Available from URL. <http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html>.
- [7] Collins J, Education Techniques for lifelong learning: Writing Multiple Choice Question for Continuing Medical Education activities and self assessment modules. Radiographic, 2006; 26: 543-51.
- [8] Dianne S, Linda C, Fritz D, Brian G, Laura H, Jolda H, et al. The standards for Educational and Psychological testing: 2009. AERA Publication, Washington DC.
- [9] Si-Mui Sim, Rashia RI, Relationship between item difficulty and discrimination indices in True/false type MCQs of a para-clinical multidisciplinary paper. Ann Accad Med Singapore. 2006; 35; 67-71.
- [10] Zubairi Am, & Kassim NLA, Classical and Research analysis of dichotomously scored reading comprehension test items. Malaysian JELTs Res 2006; 2: 1-20.
- [11] Miller MD, Linn RI, Grown Fund NE, (Ed). Measurement and assessment in Teaching. 10th ed. Upper Saddle River, NJ: Prentice Hall: 2009.
- [12] Palmer EJ, Devitt PG, Assessment of Higher order cognitive skill in undergraduate education, Modified Essay or MCQs. BMC Medical Education. 2009; 7: 49.
- [13] Gujjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality MCQs from an assessment of medical students of Ahmedabad, Gujrat. Indian J Comm. Med, 2014; 39: 17-20.
- [14] Tejindar S, Piyush G & Daljet. Principals of Medical Education. 3rd ed. 2009; 70-77.
- [15] Karelia BN, Pillai A, Vegada BN. The level of difficulty and discriminative indices and relationship between them in four response type MCQs of Pharmacology summative tests of 2nd year MBBS students. IeJSME, 2013; 6: 41-6.
- [16] Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discriminative indices of MCQs in formative exam in physiology South East Asian J Med Edu. 2013; 7: 45-50.
- [17] MCI, Regulation on Graduate Medical education (Amendment), Board of Governors in the suppression of the Medical Council of India, New Delhi 2019: 1-99.
- [18] Jonathan A, Odukoya, Olajide A, Angie O, Igbinoba, Aflobi A, Item analysis of university-wide multiple choice objective examination: the experience of a Nigerian private university, Quall Quant, 2018; 52: 983-97.
- [19] Mitra NK, Nagaraja HS, Ponnudural G & Judson JP. The level of difficulty and discriminative indices in type-A MCQs of Pre-Clinical Semester-Multidisciplinary summative tests. IJSME, 2009; 3: 2-7.
- [20] Mehta G & Makhosi V. Item analysis of MCQs-An Assessment of the assessment tool. IJHSR. 2014; 4: 197-202.
- [21] Rao C, Kashan Prasad HL, Sajitha K, Shetty J, Item analysis of MCQs: Assessing an assessment tool in medical students, I J Edu Psychology Res 2016; 2: 201-4.
- [22] Rajkumar P, Sachin BP, Kamesh V, Abhijit Vinod B, Evaluation of Multiple Choice Questions by item analysis in a medical college at Pondicherry, India, I J Community Med Public Health 2016; 3: 1612-6.
- [23] Bhat Sk, Prasad KH, Item analysis and optimizing multiple choice questions for a variable question bank in Ophthalmology: A cross sectional study, I J Ophthalmology, 2021; 69: 343-6.