International Journal of Basic & Clinical Pharmacology
Jaiswal KM et al. Int J Basic Clin Pharmacol. 2023 Jan;12(1):43-48
http://www.ijbcp.com
pISSN 2319-2003 | eISSN 2279-0780

DOI: https://dx.doi.org/10.18203/2319-2003.ijbcp20223353

Original Research Article

Post-valuation quality check of multiple-choice questions

Kavita M. Jaiswal*, Sujata Dudhgaonkar, Piyush Gharade, Nikita Sharma

Department of Pharmacology, Government Medical College, Gondia, Maharashtra, India

Received: 10 October 2022
Accepted: 04 November 2022

***Correspondence:**
Dr. Kavita M. Jaiswal,
Email: jaiswalkavita37@yahoo.com

ABSTRACT

Background: Multiple choice questions find important place in assessment in medical curriculum. Each MCQ is called item. Item analysis is quality check of MCQs after valuation of response sheets. This serves to recognize flaws in MCQs so that the given questions can be preserved in question bank, modified or deleted.
Methods: 140 medical students of second MBBS were assessed pharmacology through 20 single best response type MCQ. Post exam validation of MCQs was done by item analysis. Each item was analyzed for level of difficulty, ability of the question to discriminate between poor and good performing students and distracter effectiveness.
Results: Score '1' was given for correct and '0' for incorrect or no response. The mean score of the test was found to be 10.58±2.48, with a range 5-18. The whole test had an acceptable difficulty level with 43.25±17.81 mean difficulty index. The discrimination index of the whole test was found to be 0.123±0.184 mean PBI correlation coefficient, which is not satisfactory. Out of 60 distractors, 51 were found to be functional, hence distractor efficiency of the given test was 85%±19.96% which is acceptable.
Conclusions: 60% questions were found to be ideal and 25% were acceptable with revision of stem or options. One question was easy with poor discrimination which needs major modification in construct or subject to be placed in question bank while another one question was difficult but had negative biserial index which needs to be deleted from the question bank.

Keywords: Difficulty index, Discrimination index, distracter effectiveness, Item

INTRODUCTION

Competency based medical education (CBME) is an outcome-based education which uses competency framework to design, deliver, assess and evaluate the curriculum. Exams has important role in medical education and it enhances learning. As per CBME regular formative and summative assessment of students performance should be conducted throughout the course as a part of instruction process.[1] For theory exam assessment multiple tools are used like long answer questions (LAQs), short answer questions (SAQs) and multiple choice questions (MCQs). Quality of MCQ affects learning. Recall type MCQ encourages superficial learning. A well-constructed MCQ can assess students at high levels of Bloom's taxonomy assess sing knowledge, professional skills and advance level of critical thinking of undergraduate and post graduate medical students.[2-6] Designing such MCQs is a laborious job. The reliability of the test depends on the quality of MCQs. A single MCQ in a MCQ test paper is an item Each item consists of a stem and four options. Correct or the best answer is called key and other three are incorrect are called distracters. Quality check of MCQ is called item analysis. Each item is analyzed by four factors; how many students have attempted the given question, the difficulty level of the question, has the key been chosen by good students? and were the distracters effective? Item analysis serves two purposes. It helps to identify defective MCQ and it also helps identify the skills not mastered by students

finding the question difficult. Item analysis has three components; difficulty index-measures item difficulty, discrimination index-differentiate between good and poor performers and distractor effectiveness-is determined on basis of number of MCQs with non-functional distractors (NFD) (option selected by <5% of students) in it. Post examination analysis of the MCQs helps to assess the quality of individual test items and test as a whole. Poor items can be modified or removed from the store of questions/question bank. Hence the study was carried out with the objectives to evaluate the MCQ based on difficulty index, discrimination index and distractor effectiveness to help identify ideal MCQs in the given paper to develop a valid pool of questions and work on other questions for their inadequacies.

METHODS

The present study was carried out in department of pharmacology of a tertiary care teaching institute. A total of 140 undergraduate medical students of second year MBBS were assessed in subject of pharmacology in July 2022. It formed a part of first summative assessment consisting a 3-hour written paper with multiple choice questions (MCQs/items) to be completed in first 30 min. The number of items were 20 single best type questions. Each item consists of a stem and four options, a key and other three distracters. Students were asked to select one best answer from these four choices. Score '1' was given for each correct response and '0' for incorrect or no response, with no negative marking. Post exam validation was done by item analysis. SPSS version 24 was used for data analysis. After correction of response sheets, they were arranged in rank order highest to lowest marks. This distribution was broken in two groups i.e., high achievers (H) and lower achievers (L) based on Kelly's deviation.[7] For each item the number of students ticking option a, b, c, or d in each group was counted Each item was analyzed for three indices**.** Difficulty index represented by (DIF) or (P) assesses the difficulty level of question, ranges between 0-100% is calculated by following formula;

$$DIF\ (P) = \ (H + L)/N\ x\ 100$$

Where H is high marks achievers, L is low marks achievers, N is total number of students appearing for the paper. If P<30% the question is difficult, P=30-70% the question is acceptable and P>70% the question is easy. Higher the difficulty index lower is the difficulty of the question.

Discrimination index (D)

It measures the differences between the percentages of students in the upper group with that of the lower group who obtained the correct responses-and ranges between 0-1. The higher the discrimination index, the test item can discriminate better between students with higher test scores and those with lower test scores. The items were analyzed by calculating point- biserial correlation. One of the most accepted ways to evaluate an item is to calculate a correlation. The technical term for the correlation used in exam item analysis is a point-biserial. In a point-biserial correlation test scores on a continuous scale are compared to a single item that has only two possible values: correct or incorrect. At a high level, what you are doing is correlating response to a single question with the student's overall score. The overall test score is a signal of whether the student is high-performing or low-performing. If well written, students' responses to a given item will correlate with their overall test scores.

$$Point\ biserial\ index\ (PBI) = \ (H - L)/N\ x2$$

Normal values range between 0-1. If PBI=0-0.19-poor discrimination if PBI is between 0.2 and 0.29-acceptable discrimination. PBI between 0.3 and 0.39 is good discrimination if PBI>0.4 excellent discrimination. If PBI is negative then the item is defective or wrong key, should be discarded from question bank. The difficulty index and discrimination index are reciprocally related. Distractor effectiveness is determined on basis of number of items with non-functional distractors (NFD) (option selected by <5% of students) in it. Functional distractor (FD) is the option selected by 5% or more students. On the basis of NFDs in an item, DE ranges from 0% to 100%. Three NFDs-DE is 0, two NFD-DE is 33.3%, one NFD –DE is 66.6%, zero NFDs-DE is 100%. Ideal MCQ is a one with difficulty index (P) =30-70, discrimination index (D) > 0.24 and DE of 85.15% which is close to items having one NFD.

RESULTS

In the present study the test paper had 20 multiple choice questions (MCQs) of choose the best answer type. Each question carried 1 mark. Total 140 undergraduate students attempted the test. Score '1' was given for each correct response and '0' for incorrect or no response. The mean score of the test was found to be 10.58±2.48, with a range 5-18. Difficulty index-DIF (difficulty level) of each item is depicted in (Table 1). 60% (12 questions) were in the acceptable range having DIF between 30%-70%. Thirty percent (6 questions) were too difficult with DIF <30%. Five percent (1 question) was too easy having DIF >70%. The whole test had an acceptable difficulty level with 43.25±17.81 mean difficulty index. Discrimination index-DI (discriminating among good and poor performers in the given test) is shown in (Table 2). Fifteen percent (3 questions) had excellent high point biserial index (PBI) (>0.4). Ten percent (2 questions) had good PBI (0.3-0.39). Five percent (1 question) had acceptable PBI (0.2-0.39). Forty-five percent (9 questions) had poor PBI (<0.2). Twenty-five percent (5 questions) had negative PBI. The discrimination index of the whole test was found to be 0.123±0.184 mean PBI correlation coefficient, which is poor. Distractor effectiveness (DE) of the items is shown in (Table 3). In the present study it was found that 70% (14 questions)

had zero nonfunctional distractor (NFDs) hence DE was excellent with100% distracter efficiency.

Table 1: Difficulty index of each item (n=140).

Question no.	Correct option (key)	H	L	DIF/P	Inference about the item	Range of DIF
Q-1	b	48	21	49.28	Acceptable	For 12 questions DIF is between 30-70%
Q-3	a	46	17	45		
Q-4	a	61	33	67.14		
Q-5	b	44	27	50.71		
Q-6	d	51	30	57.85		
Q-7	a	51	15	47.14		
Q-8	b	45	33	55.71		
Q-11	a	40	46	61.14		
Q-14	a	32	25	40.71		
Q-15	b	38	40	55.71		
Q-16	d	34	28	44.28		
Q-18	a	38	37	53.57		
Q-2	b	10	12	15.71	Difficult	For 6 questions DIF is <30%
Q-10	b	22	14	25.71		
Q-12	a	17	15	22.85		
Q-17	b	13	13	18.57		
Q-19	a	10	6	11.42		
Q-20	d	20	20	28.57		
Q-13	c	43	56	70.71	Easy	For 1 question DIF is >70%
Q-9	Mistake in stem of this item					-

n-is number of students appearing, H is high marks achievers (first half scores), L is low marks achievers (second half scores), DIF/P-difficulty index. Mean DIF=43.25±17.34

Table 2: Discrimination index of each item (n=140).

Question no.	DI	Inference about the items	Point Biserial
Q3	0.41	Excellent	For 3 items >0.4
Q4	0.4		
Q 7	0.51		
Q1	0.38	Good	For 2 items 0.3-0.39
Q6	0.30		
Q5	0.24	Acceptable	For 1 item 0.2-029
Q8	0.17	Acceptable with modification	For 12 questions <0.2
Q10	0.11		
Q12	0.02		
Q14	0.10		
Q16	0.08		
Q17	0		
Q 18	0.01		
Q 19	0.05		
Q 20	0		
Q2	0.02		
Q11	0.08		
Q13	0.18		
Q15	-0.02	Deleted	Negative
Q-9	Mistake in stem of this item		

n-number of student appearing for test, DI-Discrimination index. Mean DI =0.1235±0.184.

Twenty percent (4 questions) had 1 NFD, DE was good. Five percent (1 question) had 2NFDs, DE was moderate. No question had 3 NFDs. In the given test, there were 20 questions each with 4 options. In each question one option was correct called key and rest three were incorrect or distracters. For 20 questions there were 80 options with 20 keys and 60 distractors. Out of 60 distractors, 51 were found to be functional, hence distractor efficiency of the given test was 85%±19.96% which is inacceptable range. All three indices for each

question are depicted in (Table 4). Total 6 questions (Q1, Q3, Q4, Q5, Q6, Q7) were ideal MCQs in terms of all the three indices. Three questions (Q17, Q19, Q20) had one acceptable index rest of the 10 questions had two acceptable indices out of three.

Table 3: Distractor effectiveness of each item (n=140).

Question.no	Key and Distracter (options)				DE	NFD (in circle)
	A	B	C	D		
Q2	99	22	12	7	Excellent	14 questions had zero NFDs
Q3	63	8	20	48		
Q5	33	71	17	18		
Q6	12	15	29	84		
Q7	66	42	13	19		
Q10	17	52	19	52		
Q12	39	27	50	23		
Q14	64	9	53	14		
Q15	20	101	11	8		
Q16	15	9	29	86		
Q17	35	32	46	26		
Q 18	99	19	14	8		
Q 19	19	56	25	36		
Q 20	14	17	60	49		
Q1	12	69	56	3	Good	4 questions had 1 NFD
Q4	94	29	15	2		
Q8	26	78	31	5		
Q11	112	13	6	9		
Q13	6	7	125	2	Moderate	1 question had 2 NFDs
Q9	Mistake in stem of this item					

n-is number of students appearing, DE-distractor effectiveness, number in yellow is the key, number in blue is non-functional distractor (NFD).

Table 4: Item analysis for three indices (n=140).

Item no	DIF/P	PBI	DE	Remark about item
Q1	49.28 (A)	0.38 (G)	1 NFD (G)	Ideal
Q 3	45 (A)	0.41 (E)	0 NFD (E)	
Q 4	67.14 (A)	0.4 (E)	1 NFD (G)	
Q 5	50.71 (A)	0.30 (G)	0 NFD (E)	
Q 6	57.85 (A)	0.24 (A)	0 NFD (E)	
Q 7	47.14 (A)	0.51 (E)	0 NFD (E)	
Q 8	55.71 (A)	0.17 (P)	1 NFD (G)	Acceptable
Q 11	61.14 (A)	0.08 (P)	1 NFD (G)	
Q 14	40.71 (A)	0.10 (P)	0 NFD (E)	
Q 15	55.71 (A)	0 (P)	0 NFD (E)	
Q 16	44.28 (A)	0.08 (P)	0 NFD (E)	
Q 18	53.57 (A)	0.01 (P)	0 NFD (E)	
Q 10	25.71 (D)	0.11 (P)	0 NFD (E)	Acceptable with modification
Q 12	22.85 (D)	0.02 (P)	0 NFD (E)	
Q 17	18.57 (D)	0 (P)	0 NFD (E)	
Q 19	11.42 (D)	0.05 (P)	0 NFD (E)	
Q 20	28.57 (D)	0 (P)	0 NFD (E)	
Q 13	70.71 (Ea)	0.01 (P)	2 NFD (E)	
Q 2	15.71 (D)	-0.02 (P)	0 NFD (E)	Negative biserial
Q 9	Mistake in stem of item			

DIF/P-Difficulty index, PBI-Point biserial index (discrimination index), DE-Distractor effectiveness, A-acceptable, D-difficult, P-poor, E-excellent, G-good, Ea-easy

During correction of answer sheets, it was found that question no. 9 had 3 correct options because in the stem the word 'except' was missing, that was due to printing mistake It was decided by the academic committee of the

department, that every student to be granted 1 mark for the said question.

DISCUSSION

MCQs are important part of medical exams and can assess large number of students in a short span of time. Framing MCQs as per guidelines of writing stem and options is a time-consuming laborious job. Another important aspect is evaluation of MCQs for its quality. After correction of MCQ test paper a teacher needs to know, how good the test questions were and whether the test items reflected students' performance in the course related to learning.[8] A single MCQ in a MCQ test paper is an item. Item analysis means assessing the item from students' responses for its reliability and validity by determining certain parameters like discrimination index (DI) or point biserial index (PBI), difficulty index (DIF I) and distractor efficiency (DE).[9] Item analysis of MCQs helps the examiner to recognize the inadequacies or any error met during item construction. It also aids to decide those items that are good and those that need improvement or deletion from the question bank. In the present study items/MCQs of first summative exam in subject of pharmacology were validated post examination. The mean score of the test was found to be 10.58 ±2.48, with a range 5-18. Sixty percent items were in the acceptable range having difficulty index (P) in the range of 30%-70%. Thirty percent items were too difficult with p< 30%. Five percent items were too easy having p>70%. The whole test had an acceptable difficulty level with 43.25±17.81 mean difficulty index. These findings coincide with findings from other studies. In an item analysis study conducted by Patel et al, the authors reported that 80%items were in acceptable range (p=30-70%), whereas 20% items were in unacceptable range (p<30% or p>70%).[10] Another study by Patil et al reported mean difficulty index of 48.90±13.72. Twenty - five percent items were ideal (P=50-60%), thirty-five percent items were too difficult (p<30%) whereas 18% items were too easy (p>70%).[11] Mehta et al in their study reported a mean P of 63.06±18.95 with difficulty index of 62% items in the acceptable range (P=30-70%), thirty-two percent items were too easy and 6% items were too difficult.[12] Too easy or too difficult items needs revision and should be kept for subsequent use along with items within acceptable range. Another important parameter of item analysis used for discrimination among high achievers and low achievers is discrimination index denoted by point biserial index (PBI). It has range from 0 to 1, where a greater value shows increased ability of MCQ to discriminate between a high achieving student and a low achieving student. Sometimes the value can be negative indicating a error in item and it is called negative discrimination index. This negative value shows that more number of low achievers correctly answered the question as compare to high achievers may be due to wrongly marked answer key or unclear questions. In the present study 15% questions had excellent high point biserial index (PBI) (>0.4) Ten percent questions had good PBI (0.3-0.39) and five percent had acceptable PBI (0.2-0.39). Forty-five percent questions had poor PBI (<0.2). Twenty-five percent questions had negative PBI. The discrimination index of the whole test was found to be 0.123± 0.184 mean PBI correlation coefficient, which is not satisfactory. The findings from other studies are as follows. In a study by Benish Mehmood et al discrimination index of the test was satisfactory with 0.23±0.15 mean point biserial correlation coefficient. 44% questions, that had poor point biserial index (<0.2), twenty-four percent questions had a fair PBI (0.2-0.29), twelve percent items had a good PBI (0.3-0.39) and 20% questions had an excellent PBI (0.4-0.7). [13] Another similar study conducted by Singh et al, reported 20% items with PBI≥0.20 & ≤0.35 while 30% items had <0.2 PBI.[14] Patel et al in their study reported PBI ≥0.20, and ≤0.35 for 21 items.[10] Study by Mehta et al showed mean PBI of 0.33±0.18.19.[12] A good quality item should consist of reasonable distractors. In the present study it was found that 70% questions were excellent with no non-functional distracters (NFDs) hence 100% distracter efficiency. Twenty percent questions had 1 NFD, DE was good. Five percent questions had 2NFDs, DE was moderate. No question had 3 NFDs. In the given test out of 60 distractors, 51 were found to be functional, hence distractor efficiency of the given test was 85%±19.96% which is in standard range. The DE of some studies reported in literature is as follows. In a study by Benish Mehmood et al the mean DE was found to be 85.33% ± 21.69%. Sixty-four percent items had 100% distractor efficiency where as 28% questions had one non-functional distractor while 8% MCQs had 2 non-functional distractors in each item. A study by Hingorjo et al., reported a mean DE of 81.4% while Gajjar et al reported a mean DE of 88.6±18.6 showing good efficiency of distractors.[15,16]

CONCLUSION

In the present study the whole test had an acceptable difficulty level with 43.25±17.81 mean difficulty index. Distractor efficiency was 85%±19.96% which is in a good range. Discrimination index is 0.123±0.184 mean PBI correlation coefficient, which is found to be unsatisfactory hence could not identify the poor learners. Sixty percent questions were found to be ideal or acceptable. Twenty –five percent questions were acceptable with revision of stem or options. One question was easy with poor discrimination which needs major modification in construct or subject to be placed in question bank while another one question was difficult but had negative biserial index which needs to be deleted from the question bank**.** The results of this study will initiate a change in the way MCQ test items are selected for conducting further exams and there should be a proper assessment strategy as part of the curriculum development. Much more of these kinds of analysis should be carried out after each examination to identify the areas of potential weakness in the one best answer type of MCQ tests to improve the standard of assessment

and develop a valid question bank. Such analysis will also identify poor performers and their un mastered skills.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Frank JR, Mungroo R, Ahmad Y, M Wang. Toward competency based medical education in medicine: a systemic review of published definations. Med Teacher J. 2010;32:6.
2 Stringer K, Santen SA, Lee E, Rawls M, Bailey J, Richards A, et al. Examining Bloom's taxonomy in multiple choice questions: students' approach to questions. Med Sci Edu. 2021;31:1311-7.
3 Baig M, Ali SK, Ali S, Huda N. Evaluation of multiple choice and short essay question items in basic medical sciences. Med Educ. 2007;41(8):754-62.
4 Peck C. Principles of sound assessment practice in health professions education. EC Psychol Psychiat. 2017;5(5):150-7.
5 Oyebola DO, Adewoye OE, Iyaniwura JO, Alada RA, Fasanmade AA, Raji YA. Comparative study of students'performance in preclinical physiology assessed by multiple choice and short essay questions. Afr J Med Med Sci. 2000;29(3-4):201
6 Kemp JE, Morrison GR, Ross SM. Developing evaluation instruments. New York: MacMillan College Publishing Company; 1994:180-213.
7 Kelley TL. The selection of upper and lower groups for validation of test items. J Educ Psychol. 1939;30:17-24.
8 Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The of level of difficulty and discrimination indices in type A multiple choice questions of nPre. Clin Sem Multidisclip J. 2009;3(1):2-7.
9 Considine J, Botti M, Thomas S. Design, format validity and reliability of multiple choice questions for use in nursing research and education. Collegian. 2005;12(1):19-24.
10 Patel KA, Mahajan NR. Itemized analysis of questions of multiple choice question exam. Int J Sci Res. 2013;2(2):279-80.
11 Patil VC, Patil HV. Item analysis of medicine multiple choice questions (MCQs) for under graduate (3rd year MBBS) students. Res J Pharm Biol Chem Sci. 2015;6(3):1242-51.
12 Mehta G, Mokhasi V. Item analysis of multiple choice questions-An assessment of the assessment tool. Int J Health Sci Res. 2014;4(7):197-202.
13 Benish M, Sehar S, Muhammad A, Asma S, Nauman BK, Zainab N. Quality analysis of multiple choice questions. Pak Oral Dent J. 2020;4(40):235-9.
14 Singh JP, Kariwal P, Gupta SB, Shrotriya VP. Improving multiple choice questions through item analysis: an assessment of the assessment tool. Int J Basic Sci Appl Res. 2014;1(2):53-7.
15 Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc. 2012;62(2):142-7.
16 Gajjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality multiple choice questions (mcqs) from an assessment of medical students of Ahmedabad, Gujarat. Indian J Commun Med. 2014;39(1):17-20.

Cite this article as: Jaiswal KM, Dudhgaonkar S, Gharade P, Sharma N. Post-valuation quality check of multiple-choice questions. Int J Basic Clin Pharmacol 2023;12:43-8.