Oncology
and Radiotherapy ©
Vol.14 Iss.5:012-016 • RESEARCH ARTICLE

Assessment of MCQs in MBBS program in Saudi Arabia

Omer Abdelgadir Elfaki[1], Abdulaziz Alamri[2], Karim Eldin M A Salih[3]

[1]Department of Medical Education, King Khalid University and Batterjee Medical College, Jeddah, Saudi Arabia
[2]Department of Medical Education and Consultant Urologist, College of Medicine, KKU, Abha, Saudi Arabia
[3]Department of Pediatrics, Medical Education, College of Medicine, University of Bisha, Saudi Arabia

SUMMARY

Background: Multiple choice questions represent one of the commonest methods of assessment in medical education. They believed to be reliable and efficient. Their quality depends on good item construction. Item analysis is used to assess their quality by computing difficulty index, discrimination index, distractor efficiency and test reliability.

Objective: The aim of this study was to evaluate the quality of MCQs used in the College of Medicine, King Khalid University, Saudi Arabia.

Design: Cross sectional Study design

Materials and Methods: Item analysis data of 21 MCQs exams were collected. Values for difficulty index, discrimination index, distractor efficiency and reliability coefficient. Descriptive statistic parameters were computed.

Results: Twenty-one tests were analysed. Overall, 7% of the items among all the tests were difficult, 35% were easy and 58% were acceptable. The mean difficulty of all the tests was in the acceptable range of 0.3-0.85. Items with acceptable discrimination index among all tests were 39%-98%. Negatively discriminating items were identified in all tests except one. All distractors were functioning in 5%-48%. The mean functioning distractors ranged from 0.77 to 2.25. The KR-20 scores lie between 0.47 and 0.97.

Key words: assessment, MCQ, tests

Address for correspondence:

Omer Abdelgadir Elfaki, Prince Abdullah Alfaysal Batterjee Medical College, North Obhour Jeddah 21442 P O Box 6231, KSA, Saudi Arabia, email: elfaroog@gmail.com

Word count: 2817 **Tables:** 05 **Figures:** 00 **References:** 21

Received: - 02 September, 2020

Accepted: - 24 September, 2020

Published: - 30 September, 2020

INTRODUCTION

There is wide agreement that the ultimate goal of undergraduate and postgraduate medical education is improvement of health of individuals and community [1]. Therefore, medical education programs strive to design their outcomes to that end and align student assessment to ensure adequate competency. One of the commonest methods of assessment in medical education is Multiple Choice Questions (MCQs) [2]. They are proved to be reliable, efficient and fair but that is not inherited in them and is dependent on many factors among which are adequate psychometric properties and good construction [3]. Item analysis is quite useful for assessment of quality of MCQs items and tests [4]. This helps in revision and improvement of these items and tests [5]. The quality parameters identified by item analysis include Difficulty Index (D), Discrimination Index (DIS), Distractor Efficiency (DE) and test reliability. The D is the percentage of students who answered an item correctly and ranges between 0 and 100% [6, 7]. DIS describes the ability of an item to differentiate between higher and lower ability students and ranges between -1 and 1 [8]. For distractor function analysis, any distractor selected by less than 5% of the examinees is considered Non-Functional Distractors (NFD). The KR20 formula was used to measure reliability of each test. Values ≥ 0.80 are considered acceptable [9]. Thus, item analysis provides valuable information about the reliability and validity of a test item. The objective of the present study was to assess the quality of MCQs used in the summative exams at the end of semester in the College of Medicine, King Khalid University, Saudi Arabia.

MATERIALS AND METHODS

Settings

This cross-sectional study was conducted in the department of medical education of the College of Medicine, King Khalid University, in the period January-February 2017. The college was established in 1980 as the first one in the southern region of Saudi Arabia. The MBBS program in the college is seven years long including the internship. The curriculum is discipline-based and is divided into preclinical phase and a clerkship phase. The preclinical phase is devoted for basic sciences. Male and female students are taught the same curriculum separately in

two different campuses. Exactly the same assessment is applied to all students in the two campuses. One best answer MCQs are used through all subjects and disciplines for both formative and summative purposes. They form an important component with different weights from the final assessment.

Data collection

The item analysis of MCQs tests for 21 courses at the end of first semester for the academic year 2016-2017 were collected. All of the tests were summative tests developed by the course instructors and approved by the relevant departments. All of the tests were also screened for item-writing flaws by assigned experts in MCQs construction before final approval. The studied quality parameters of item analysis included three for items and one for the whole test. Those which assess the quality of the items were the D, DIS and DE. The D and DIS were calculated according to the following formulas.

$D = [(U + L)/N] \times 100$

$DIS = 2 \times [(U – L)/N]$

Where U represent the upper 27% students, L the lower 27% and N the number of students.

Data analysis

All data was entered into MS Excel 2013 and descriptive statistics parameters were computed. Items with D values below 30%, more than 85% and between 30%-85% were classified as difficult, easy and acceptable respectively [8]. Regarding DIS, in this study, items were categorized into four groups based on their DIS values: below zero, zero, >0-<0.2 and ≥ 0.2. Only those with DIS values ≥ 0.2 were considered acceptable. Items in each test were counted to have one, two, three or nil NFDs based on the criteria for NFD as that selected by less than 5% of the examinees. For the tests, reliability measured by KR-20 was assessed. Ethical approval for this study was obtained from the research ethics committee of the college of medicine, King Khalid University.

RESULTS

The number of MCQs tests studied was 21, covering all levels of study from year two to six. The number of items in each test ranged from 25 to 87 (Table 1).

The percentage of difficult items ranged from zero to 36%. Seven to 68% and 26% to 84% of items were classified as easy

Tab. 1. Basic descriptive data

Description	Test										
	1	2	3	4	5	6	7	8	9	10	11
Study level (year)	4	2	2	3	3	6	4	3	4	2	6
No of items	25	80	30	80	60	60	60	60	80	87	60
Description	Test										
	12	13	14	15	16	17	18	19	20	21	
Study level (year)	5	2	2	5	5	3	4	4	6	4	
No of items	39	73	30	25	50	76	70	40	60	60	

Tab. 2. Distribution of MCQs among the difficulty index groups

Test	Difficult (%) <30%	Easy (%) >85%	Acceptable (%) 30%-85%	Number of items	Mean difficulty index
1	9 (36)	7 (28)	9 (36)	25	0.51
2	5 (6)	9 (11)	66 (83)	80	0.62
3	0	14 (47)	16 (53)	30	0.76
4	2 (3)	10 (13)	68 (84)	80	0.6
5	11 (18)	4 (7)	45 (75)	60	0.56
6	4 (7)	20 (33)	36 (60)	60	0.71
7	2 (3)	27 (45)	31 (52)	60	0.78
8	0	16 (27)	44 (73)	60	0.66
9	6 (8)	14 (17)	60 (75)	80	0.54
10	4 (5)	47 (54)	36 (41)	87	0.79
11	3 (6)	41 (68)	16 (26)	60	0.85
12	5 (12)	15 (39)	19 (49)	39	0.78
13	7 (10)	23 (31)	43 (59)	73	0.66
14	4 (14)	8 (28)	18 (58)	30	0.66
15	2 (8)	14 (56)	9 (36)	25	0.75
16	4 (7)	24 (48)	22 (45)	50	0.66
17	2 (3)	21 (27)	53 (70)	76	0.68
18	6 (8)	37 (53)	27 (39)	70	0.75
19	5 (13)	5 (12)	30 (75)	40	0.55
20	4 (6)	35 (59)	21 (35)	60	0.81
21	3 (4)	11 (19)	46 (77)	60	0.66
Total	74 (7)	386 (35)	640 (58)	1100	

Tab. 3. Distribution of MCQs among the discrimination index values

Test	% of Negative values	% of zero values	% of values in the range 0-0.19	% of values in the range 0.2 or above	Mean discrimination index
1	12	4	4	80	0.64
2	3	10	0	87	0.56
3	0	6	12	82	0.46
4	1	1	0	98	0.6
5	12	8	0	80	0.6
6	15	38	1	46	0.12
7	6	24	9	61	0.29
8	2	3	9	86	0.42
9	5	8	8	79	0.4
10	5	56	0	39	0.2
11	3	37	16	44	0.27
12	11	34	5	50	0.2
13	4	40	1	54	0.36
14	17	27	3	53	0.25
15	4	6	37	53	0.2
16	10	30	0	60	0.2
17	2	3	5	90	0.5
18	8	40	2	50	0.21
19	10	12	6	72	0.32
20	13	28	15	44	0.19
21	5	13	8	74	0.38

Tab. 4. Distractor analysis of MCQs items in all tests

Test	Three NFDs	Two NFDs	One NFD	Nil	No of questions	Functioning distractors per test	Mean functioning distractors per item
1	38%	27%	16%	19%	25	39%	1.17
2	5%	25%	40%	30%	80	65%	1.95
3	22%	33%	23%	22%	30	48%	1.44
4	12%	37%	32%	19%	80	52%	1.56
5	5%	27%	47%	21%	60	61%	1.83
6	22%	30%	33%	15%	60	47%	1.41
7	18%	35%	35%	12%	60	47%	1.41
8	10%	30%	40%	20%	60	57%	1.71
9	9%	16%	30%	45%	80	70%	2.11
10	32%	30%	26%	12%	87	39%	1.17
11	52%	27%	15%	6%	60	26%	0.77
12	18%	40%	33%	9%	39	44%	1.32
13	14%	22%	36%	28%	73	59%	1.77
14	12%	30%	37%	21%	30	56%	1.68
15	44%	11%	28%	17%	25	39%	1.17
16	38%	30%	26%	6%	50	33%	1
17	3%	16%	34%	47%	76	75%	2.25
18	17%	21%	10%	52%	70	65%	1.94
19	3%	18%	31%	48%	40	75%	2.25
20	43%	28%	22%	7%	60	31%	0.93
21	8%	22%	37%	33%	60	65%	1.95

and acceptable respectively. Overall, 7% of the items among all the tests were difficult, 35% were easy and 58% were acceptable (Table 2).

Analysis of the DIS values revealed all the tests contained some items which were negatively discriminating except one (Table 3). In eight exams, items with negative DIS reached 10% or more, while in seven exams 80% or more of the questions had acceptable DIS. The range of percentage of items with acceptable DIS values in all tests were 39%-98%.

The percentage of items with NFDs are shown. Seven percent to 75% of distracters were functional in all of the exams. All the distracters were functional in 5%-48% of the items in all exams (Table 4).

Means of D, DIS, FD per item and KR-20 scores are shown in Table 5.

Tab. 5. Summary of item analysis parameters

Test	No of items	Mean D	Mean DIS	Mean FD per item	KR-20
1	25	0.51	0.64	1.17	0.57
2	80	0.62	0.56	1.95	0.95
3	30	0.76	0.46	1.44	0.93
4	80	0.6	0.6	1.56	0.97
5	60	0.56	0.6	1.83	0.89
6	60	0.71	0.12	1.41	0.83
7	60	0.78	0.29	1.41	0.82
8	60	0.66	0.42	1.71	0.9
9	80	0.54	0.4	2.11	0.89
10	87	0.79	0.2	1.17	0.72
11	60	0.85	0.27	0.77	0.72
12	39	0.78	0.2	1.32	0.51
13	73	0.66	0.36	1.77	0.83
14	30	0.66	0.25	1.68	0.85
15	25	0.75	0.2	1.17	0.47
16	50	0.66	0.2	1	0.61
17	76	0.68	0.5	2.25	0.92
18	70	0.75	0.21	1.94	0.77
19	40	0.55	0.32	2.25	0.85
20	60	0.81	0.19	0.93	0.68
21	60	0.66	0.38	1.95	0.85

DISCUSSION

This was a cross sectional descriptive study in a medical school where a discipline-based curriculum had been implemented. Twenty-one MCQs final exams were analysed representing 21 different courses at all study levels. This was the first study done in the college for the same purpose and could be the first one covering all the MBBS program in a Saudi medical school context as far as the authors know. The mean D of all the tests was in the acceptable range of 0.3-0.85. In fact, all of the means were in the range of 0.51-0.85 indicating that the items tended to be more near the easy end. This is also shown by the percentage of easy questions among all the tests which reached 35% compared to 7% difficult questions. In a similar study where 12 tests were analysed, the D scores ranged from 64% to 89% [10]. In another study, 40% of the items had D values of more than 70% and were classified as easy [11]. It was also identified that in five exams, the easy questions reached more than 50%. Overall, the mean of DIS values were acceptable. Only two of the exams, number 6 and 20, had unacceptable mean DIS of less than 0.2. However, all the exams contained negatively discriminating questions except exam number 3. In exam 14, the negatively discriminating items were 17%. In similar studies but analysing lesser number of items, 20% (12) and 4% (2) of the items were negatively discriminating [12]. The percentage of NFDs ranged between 3% and 52% among all exams. This could explain the high D noticed in some of the exams. It seemed that the distracters used in those questions were not plausible. Haladyna and Downing concluded that more than 38% of distractors on the tests were NFDs and were eliminated [13]. Tests become more valid and reliable whenever the number of plausible distractors increases [14]. A NFD in an item lowers the quality of that item and should be revised or removed [15]. It had been found in some studies that items with three functioning distractors ranged from 1.1 to 8.4% of all items. Teachers might expect up to 50% of the items they develop not to perform as they expect [16]. Distractors chosen by students can be helpful in identification of the learning difficulties experienced by them [17]. The mean number of functioning distractors per item ranged from 1.35 to 1.74 in one of the studies where 514 items with three options were analysed. Our findings identified values of mean functioning distractors in the two extremes in the range of 0.77 in test 11 to 2.25 in tests 17 and 19. In 76% of the exams the reliability was acceptable reaching above 0.7. In 62% of the exams it was more than 0.8. A reliability of less than 0.5 is considered questionable, [18] while that from 0.5 to 0.6 makes it necessary to revise the test. In the present study, only one exam has got a reliability of less than 0.5. A possible explanation for that could be the relatively low number of questions which was 25. The test scores obtained from such tests may not reflect student competency but reflect the effects of factors related to the testing conditions and peculiarities of the items [19]. The analysis of ten pharmacology MCQs exams resulted in reliability coefficient scores of 0.38 to 0.66 [20] which was below the 0.7 limit taken as the lowest acceptable reliability for a class room test [21].

CONCLUSION

Overall, the quality of the items and tests was found to be acceptable. Some items were identified to be problematic and need to be revised. The quality of few tests of specific courses was questionable. These tests need to be revised and steps taken to improve this situation.

ACKNOWLEDGEMENT

We are thankful to Mr. Muhammad Abid Khan for Data Analysis.

REFERENCES

1. Chandratilake M, Davis M, Ponnamperuma G. Evaluating and designing assessments for medical education: the utility formula. Intern J Med Edu. 2009;1:1-7.
2. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distracter efficiency. J Pak Med Assoc. 2012;62:142-146.
3. McCourbie P. Improving the fairness of multiple-choice questions: a literature review. Med Teach. 2004;26:709-712.
4. Shad M. Item Analysis of MCQs of a Pharmacology Term Exam in a Private Medical College of Pakistan. PJMHS. 2018;12:700-703.
5. Matlock-Hetzel S. Presented at annual meeting of the Southwest Educational Research Association, Austin. 1997.
6. Eaves S, Erford B. The Gale group. The purpose of item analysis, item difficulty, discrimination index, characteristic curve. 2013.
7. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Med Educ. 2009;9:1-8.
8. Ananthakrishnan N. Item analysis-validation and banking of MCQs. In: Ananthakrishnan N, Sethuraman KR, Kumar S, editors. Medical Education principles and practice. 2nd ed. JIPMER, Pondicherry: 131-137.
9. El-Uri FI, Malas N. Analysis of use of a single best answer format in an undergraduate medical examination. Qatar Med J. 2013;2013:3-6.
10. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1, multidisciplinary summative tests. Int J Med Educ. 2009;3:2-7.
11. Si-Mui S, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singapore. 2006;35:67-71.
12. 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. Ind J Commun Med. 2014;39:17.
13. Haladyna TM, Downing SM. How many options is enough for a multiple-choice test item. Educ Psychol Meas. 1993;53:999-1010.
14. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ. 2002;15:309-334.
15. Linn RL, Gronlund NE. Measurement and assessment in teaching (Eighth Edition). Upper Saddle River. 2000.
16. Haladyna TM. Developing and validating multiple-choice test items. Mahwah NJ: Lawrence Erlbaum Associate. 2004.
17. Nitko AJ, Brookhart SM. Educational assessments of students (4th ed.) Englewood Cliffs, NJ: Prentice Hall. 2004.
18. http://www.fcit.usf.edu/assessment/basic/basic.html
19. http://www.washington.edu/oea/service/scanningscoring/scanning/itemanalysis
20. Vegada BN, Karelia BN, Pillai A. Reliability of four-response type multiple choice questions of pharmacology summative tests of II MBBS students. Intern J Math Stat Invent. 2014;2:6-10.
21. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol. 1993;78;98-104.8.