

Comparison in the quality of distractors in three and four options type of multiple choice questions

Nourelhouda A A Rahma¹
 Mahdi M A Shamad²
 Muawia E A Idris¹
 Omer Abdelgadir Elfaki³
 Walyedldin E M Elfakey⁴⁻⁶
 Karimeldin M A Salih^{3,7}

¹Pediatrics Department, University of Bahri, College of Medicine, Sudan;

²Dermatology and Venereology Department, University of Bahri, College of Medicine, Khartoum, Sudan;

³Medical Education Department, King Khalid University, College of Medicine, Kingdom of Saudi Arabia;

⁴Department of Pediatrics, University of Bahri, College of Medicine, ⁵Medical Education Department, ⁶Pediatrics Department, Albaha University, College of Medicine, Kingdom of Saudi Arabia; ⁷Medical Education Department, University of Bahri, College of Medicine, Pediatrics, Khartoum, Sudan

Introduction: The number of distractors needed for high quality multiple choice questions (MCQs) will be determined by many factors. These include firstly whether English language is their mother tongue or a foreign language; secondly whether the instructors who construct the questions are experts or not; thirdly the time spent on constructing the options is also an important factor. It has been observed by Tarrant et al that more time is often spent on constructing questions than on tailoring sound, reliable, and valid distractors.

Objectives: Firstly, to investigate the effects of reducing the number of options on psychometric properties of the item. Secondly, to determine the frequency of functioning distractors among three or four options in the MCQs examination of the dermatology course in University of Bahri, College of Medicine.

Materials and methods: This is an experimental study which was performed by means of a dermatology exam, MCQs type. Forty MCQs, with one correct answer for each question were constructed. Two sets of this exam paper were prepared: in the first one, four options were given, including one key answer and three distractors. In the second set, one of the three distractors was deleted randomly, and the sequence of the questions was kept in the same order. Any distracter chosen by less than 5% of the students was regarded as non-functioning. Kuder-Richardson Formula 20 (Kr-20) measures the internal consistency and reliability of an examination with an acceptable range 0.8–1.0. Chi square test was used to compare the distractors in the two exams.

Results: A significant difference was observed in discrimination and difficulty indexes for both sets of MCQs. More distractors were non-functional for set one (of four options), but slightly more reliable. The reliability (Kr-20) was slightly higher for set one (of four options). The average marks in option three and four were 34.163 and 33.140, respectively.

Conclusion: Compared to set 1 (four options), set 2 (of three options) was more discriminating and associated with low difficulty index but its reliability was low.

Keywords: option, distractors, reliability, discrimination, MCQs

Introduction

In order to ensure the safety of human beings, accountable medical graduates are important and their competence should be certified; therefore assessment should be valid and reliable, a process which is not an easy job to perform.¹ Despite reliability, validity, coverage of a large area of the curriculum and their use for large numbers of students, multiple choice questions (MCQs) could be used to assess all domains of learning, however, MCQs need to be set by an expert, otherwise they might not serve the purpose of assessment properly.^{2,3} Many authors recommend revisions of MCQs before the exam in order to attain the course objectives.¹ With regard to the selection

Correspondence: Karimeldin Mohamed Ali Salih

Pediatrics and Medical Department, College of Medicine, University of Bahri
 Tel +966 5 4334 7641

Email karimeldin_salih@hotmail.com

of the best answer, usually one option is the correct answer whereas the other three or four options are there to serve as distractors.⁴ Well constructed MCQs can really differentiate between those who know and those who do not know; however, a lot of time and effort is needed to design well constructed MCQs.^{5,6} The numbers needed for high quality MCQs distractors will be determined by many factors. These include firstly whether English language is the candidates' mother tongue or a foreign language; secondly, whether those who construct the questions are experts or not. It has been observed by Tarrant et al that more time is often spent on constructing questions than on tailoring sound, reliable, and valid distractors.^{4,7} The distractors are considered as not distracting or doing their presumed job if they are not selected at all by examinees or only used by less than 5% of them.^{4,7} A distracter that is not functioning and might have been added just to complete the requested requisite options should be removed, and as reported by Haladyna and Downing, 38% of test distractors were excluded since they were selected by less than 5% of the examinees.⁸ Although many institutes adopted three or four distractors in addition to the key answer, the issue of non-functioning distractors raises the question of how many types of best or single best answer should be used, that of three, four or five options.⁴ The reliability for both sets of MCQs is nearly the same in this study, which in agreement with other studies.⁹ In fact, the choice of many distractors allows covering of a large amount of content, however, few distractors minimize the chance of just adding irrelevant, invalid distractors, particularly for those for whom English is a foreign language, and psychometric analysis proves there is no significant difference between the three or four types of distractors.⁸⁻¹⁰ After extensive work regarding the number of options, many authors recommend that at least three options be used.^{11,12} Item analysis is the process by which reliability, difficulty, and discrimination ability of the item are studied. All psychometric properties in 5-option MCQs were significantly affected if three of the options were not functioning while psychometric properties of items with 3 functioning options were found to be similar to those of 5-option items.¹³ In a review of literature on the number of options of MCQs, it was concluded that tests with 3-option MCQs proved to be of similar quality to those of 4- or 5-option MCQs and the recommendation was to use 3 options.¹⁴

The rationale for the study

1) MCQs is a very popular type of assessment which is sometimes considered as easy tools for assessment. 2) Proper evaluation of the efficiency of the distractors will help the institute to remove non-functioning distractors. 3) Determination of the

number of options will allow good distribution of the questions in the curriculum and will save the time of the examinees. 4) No previous study was carried out in the University of Bahri to determine non-functioning distractors.

Objectives

1) To investigate the effects of reducing the number of options on psychometric properties of the item. 2) To determine the frequency of functioning distractors among three or four options in the MCQs exam of the dermatology course in University of Bahri, College of Medicine. 3) To adopt high quality MCQs exams.

Materials and methods

This is an experimental study that had been conducted at the College of Medicine, University of Bahri (Sudan) in the period June to October 2016. University of Bahri was founded in Khartoum, Sudan in 2011 on the background of the three universities sited before in southern Sudan. Resources in developing countries at large, and Sudan specifically, were limited. The curriculum can be described as hybrid adopting SPICES model.^{15,16} The duration of medical study was six years excluding the internship. The study was performed on the MCQs exam in dermatology. The dermatology course is taught over ten weeks at the level of year five. Students have to pass the course before they become eligible for the MBBS award. Forty MCQs, of the one best answer type, were constructed by subject experts and reviewed by the exam committee before approval. This carried 30% of the weight of the final mark. Two sets of this exam were prepared: the first one was with four options: one key answer and three distractors. The three distractors were reduced randomly to two to form the second set and the sequence of the questions was kept in the same order. The exam was scheduled ten weeks before. The students initially sat the exam with four options (set 1) which continued for one hour. The students were then told to choose if they wanted to sit the second exam (set 2) immediately after the first for the same duration for research purposes. All students who sat for the four options exam agreed to sit the three options exam. Item analysis was performed for the two exams. SPSS was used for analysis. Paired Student's *t*-test was computed to compare the difficulty indices of the two exams and the discrimination indices. Distracter function was analyzed for the two exams. Any distracter chosen by less than 5% of the students was regarded as non-functioning. Kuder-Richardson Formula 20 (Kr-20) measures the internal consistency and reliability of an examination with acceptable range 0.8–1.0.¹⁷ Chi square test was used to compare the distractors in the two exams. Ethical clearance was obtained

from the Research and Ethics Committee of the College of Medicine, University of Bahri. Verbal consent was obtained from all students who agreed to participate.

Results

The results showed the mean for discrimination index for the three options (set 2) and four options (set 1) was 0.260 and 0.133, respectively, where standard deviation (S.D) was 0.133 and 0.155, respectively. The mean for difficulty index for three options (set 2) and four options (set 1) was 0.85 and 0.83, respectively where S.D was 0.130 and 0.155, respectively (Table 1). The reliability (Kr-20) for three (set 2) and four options (set 1), was more or less similar to each other 0.82 and 0.83, respectively, Table 1. Significant differences were observed in difficulty index between set 1 and 2 with a *P* value of 0.001, but not in discrimination index between three and four options with a *P* value of 0.038, 0.001 (Tables 2 and 3). The average marks in three and four options were

Table 1 Discrimination index for 3 and 4 options test, together with Kr-20 values

Discrimination index		
	3 options test	4 options test
Mean	0.260	0.280
S.D	0.133	0.155
Kr-20	0.820	0.830
Difficulty index		
	3 options test	4 options test
Mean	0.85	0.83
S.D	0.13	0.15
Kr-20	0.82	0.83

Note: Bold data indicates acceptable reliability.

Abbreviations: S.D., standard deviation; Kr-20, Kuder-Richardson Formula 20.

Table 2 Discrimination index

Option 3 & Option 4	Discrimination index				t	df	Sig. (2-tailed)	
	Mean	S.D	Std. error mean	95% confidence interval of the difference				
				Lower	Upper			
		0.02500	0.07841	0.01240	0.05008	0.00008	-2.816	39

Abbreviations: S.D., standard deviation; Std., standard; df, degrees of freedom.

Table 3 Difficulty index

Option 3 & Option 4	Difficulty index				t	df	Sig. (2-tailed)	
	Mean	S.D.	Std. error mean	95% confidence interval of the difference				
				Lower				Upper
	0.02488	0.04255	0.00673	0.01127	0.03849	3.698	39	0.001
			3 options		4 options			

Abbreviations: S.D., standard deviation; Std., standard; Sig., significance; df, degrees of freedom.

Table 4 Average mark in both sets of MCQs

Set	Average mark	P-value
Set 1	34.163	0.0001
Set 2	33.140	

Abbreviation: MCQs, multiple choice questions.

34.163 and 33.140, respectively, with a significant *P* value of 0.0001 (Table 4). Only 25% of the distractors were functioning in the three options (set 2) in contrast to 5% functioning distractors in the four options (set 1) (Table 5).

Discussion

Significant differences were observed in the difficulty index for both sets of MCQs, but no differences in discrimination. More distractors were found to be non-functioning for set 1 MCQs but they were slightly more reliable. The average marks in three and four options were 34.163 and 33.140, respectively. Our study showed that only 25% (for set 2), and 5% (for set 1) of the distractors were functioning well. Most studies agree that in most of the items only two distractors are functioning.^{8,6} It is very difficult, even for a well-trained instructor, to provide many functioning distractors, otherwise he would just add distractors for the purpose of completion; moreover, it would be quantity issues rather than quality issues since students will not select unsound additional distractors.⁶⁻⁸ It was very obvious that an item with two distractors usually provides more reasonable distractors since more options might give rise to more non-functional distractors.^{8,12} The reliability for both sets of MCQs is nearly the same in this study, which in agreement with other studies.¹⁸ In fact, there is no golden rule for the correct number of items within any exam but the minimum accepted number of options should

Table 5 Details regarding functioning distractors index

Distracter Index						
Number of functioning distractors	3 options		4 options			P-values
	Frequency	%		Frequency	%	
0 distractors	10	25	0 distractors	2	5	0.0391
1 distractors	16	40	1 distractors	13	33	0.7100
2 distractors	14	35	2 distractors	10	25	0.5410
3 distractors	N.A	N.A	3 distractors	15	38	N.A
Total	40 questions		Total	40 questions		

Abbreviation: N.A, not applicable.

not be less than three.^{12,19} Usually teachers use four option distractors either because it is the traditional practice which they have been accustomed to, or it is the practice adopted by their college, or they feel it covers a large amount of content in the curriculum.⁷ Non-functioning distractors that are easily eliminated by students lead to easy options and causes confusion with the standard settings.²⁰ In this study, significant differences in difficulty, but not discrimination indexes between set 1 and set 2 MCQs were observed. Three options (set 2) MCQs were discriminating between upper and lower achievers. The questions in set 2 were observed to be more difficult by the students. These findings are in agreement with those of other studies by Tarrant et al and Trevisan et al,^{1,4,21} but they disagree with Owen and Froman.⁹ The reliability for both sets of MCQs is nearly the same in this study. These findings call for the urgent need for proper construction of MCQs, adequate faculty training, and revision of MCQs by exam committees as well before the exam is due.²² There is no doubt that more distractors, if well designed, will add to the reliability of the test. The slight increase in reliability in the four options (set 1) type in the current study is in agreement with Tarrant et al.⁴ Some authorities believe that the choice of three options (set 2) allows teachers to construct high quality MCQs, and saves more time for teachers and students. Aamodt and McShane observed that the spared time could be used for constructing more MCQs of the three options type to cover more content.²³ There are small differences in the average marks between the two groups. Even though the average mark was higher in the three options type, it was not significant.²³

Strengths

To our knowledge, this is the first type of study to be carried out in the Sudanese educational institutes in general or in the University of Bahri, in particular.

Limitations

Our findings cannot be generalized since this study involves only one exam. The sample was not taken randomly among

other subjects. Another limitation is that the experiments of both set 1 and 2 MCQs applied to the same group rather than two groups. The use of two groups would avoid some contaminations.

Conclusion

Compared to set 1 (four options), set 2 (three options) was found to be more discriminating and associated with low difficulty index but its reliability was low.

Recommendations

We recommend the use of both sets 1 and 2 to utilize the advantages of both in the same test. Moreover, we hope all institutes at the national and international levels will conduct psychometric analysis and revise the validity of their questions. Writing high quality MCQs needs good experience and regular faculty developments.

Acknowledgments

The cooperation of our students, colleagues in the college, the clerk, and the statistician is highly appreciated. I would like to acknowledge my colleague Dr. Mohamed salah el Din for editing the final draft of this study.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Salih KM, Alshehri MA, Elfaki OA. A Comparison Between Students' Performance In Multiple Choice and Modified Essay Questions in the MBBS Pediatrics Examination at the College of Medicine, King Khalid University, KSA. *Journal of Education and Practice*. 2016;7(10):116–120.
2. Moeen-Uz-Zafar, Badr-Aljarallah. Evaluation of mini-essay questions (MEQ) and multiple choice questions (MCQ) as a tool for assessing the cognitive skills of undergraduate students at the Department of Medicine. *Int J Health Sci (Qassim)*. 2011;5(2 Suppl 1):43–44.
3. Anbar M. Comparing assessments of students' knowledge by computerized open-ended and multiple choice tests. *Acad Med*. 1991;66(7):420–422.
4. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distracters in multiple-choice questions: a descriptive analysis. *BMC Med Educ*. 2009;9:40.

5. Farley JK. The multiple-choice test: writing the questions. *Nurse Educ.* 1989;14(6):10–12.
6. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ.* 2004;38(9):974–979.
7. Haladyna TM, Downing SM. Validity of a taxonomy of multiple-choice item-writing rules. *Appl Meas Educ.* 1989;2(1):51–78.
8. Haladyna TM, Downing SM. How many options is enough for a multiple-choice test item? *Educ Psychol Meas.* 1993;53(4):999–1010.
9. Owen SV, Froman RD. What's wrong with three-option multiple choice items? *Educ Psychol Meas.* 1987;47(2):513–522.
10. Rogers WT, Harley D. An empirical comparison of three- and four-choice items and tests: susceptibility to testwiseness and internal consistency reliability. *Educ Psychol Meas.* 1999;59(2):234–247.
11. Elfaki OA, Bahamdan KA, Al-Humayed S. Evaluating the quality of multiple-choice questions used for final exams at the Department of Internal Medicine, College of Medicine, King Khalid University. *Sudan Medical Monitor.* 2015;10(4):123–127.
12. Rodriguez MC. Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educ Meas Issues Pract.* 2005;24(2):3–13.
13. Deepak KK, Al-Umran KU, Al-Sheikh MH, Dkoli BV, Al-Rubaish A. Psychometrics of Multiple Choice Questions with Non-Functioning Distracters: Implications to Medical Education. *Indian J Physiol Pharmacol.* 2015;59(4):428–435.
14. Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. *Natl Med J India.* 2008;21(3):130–133.
15. Harden RM. Ten questions to ask when planning a course or curriculum. *Med Educ.* 1986;20(4):356–365.
16. Al-Eraky MM. Curriculum Navigator: aspiring towards a comprehensive package for curriculum planning. *Med Teach.* 2012;34(9):724–732.
17. El-Uri FI, Malas N. Analysis of use of a single best answer format in an undergraduate medical examination. *Qatar Med J.* 2013;2013(1):3–6.
18. Osterlind SJ. *Constructing test items: Multiple-choice, constructed-response, performance, and other formats.* Boston: Kluwer Academic Publishers; 1998.
19. Frary RB. More multiple-choice item writing do's and don'ts. *Practical Assessment, Research & Evaluation.* 1995;4(11).
20. Case SM, Swanson DB. *Constructing Written Test Questions for the Basic and Clinical Sciences.* Philadelphia, PA: National Board of Medical Examiners; 2001.
21. Trevisan MS, Sax G, Michael WB. The effects of the number of options per item and student ability on test validity and reliability. *Educ Psychol Meas.* 1991;51(4):829–837.
22. Wallach PM, Crespo LM, Holtzman KZ, Galbraith RM, Swanson DB. Use of a committee review process to improve the quality of course examinations. *Adv Health Sci Educ Pract.* 2006;11(1):61–68.
23. Aamodt MG, McShane T. A meta-analytic investigation of the effect of various test item characteristics on test scores. *Public Personnel Management.* 1992;21(2):151–160.

Advances in Medical Education and Practice

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

Dovepress

including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.