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 Salih5,7

1Pediatrics Department, University of Bahri, College of Medicine, Sudan; 2Dermatology and Venerology 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, Sudan; 5Medical Education Department, Alba University, College of Medicine, Kingdom of Saudi Arabia; 6Medical 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
of the best answer, usually one option is the correct answer whereas the other three or four options are there to serve as distractors. 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. 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. 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. A distractor 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. 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. The reliability for both sets of MCQs is nearly the same in this study, which in agreement with other studies. 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. After extensive work regarding the number of options, many authors recommend that at least three options be used. Item analysis is the process by which reliability, difficulty, and discrimination ability of the item are is 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. 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.

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. 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. Distractor function was analyzed for the two exams. Any distractor 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. 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 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 1 Discrimination index for 3 and 4 options test, together with Kr-20 values

<table>
<thead>
<tr>
<th>Discrimination index</th>
<th>3 options test</th>
<th>4 options test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.260</td>
<td>0.280</td>
</tr>
<tr>
<td>S.D</td>
<td>0.133</td>
<td>0.155</td>
</tr>
<tr>
<td>Kr-20</td>
<td>0.820</td>
<td>0.830</td>
</tr>
</tbody>
</table>

**Note:** Bold data indicates acceptable reliability.

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

### Table 2 Difficulty index

<table>
<thead>
<tr>
<th>Option 3 &amp; Option 4</th>
<th>Discrimination index</th>
<th>Std. error</th>
<th>95% confidence interval of the difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>S.D</td>
<td>Std. mean</td>
<td>Lower</td>
</tr>
<tr>
<td>0.02500</td>
<td>0.07841</td>
<td>0.01240</td>
<td>0.05008</td>
</tr>
</tbody>
</table>

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

### Table 3 Difficulty index

<table>
<thead>
<tr>
<th>Option 3 &amp; Option 4</th>
<th>Difficulty index</th>
<th>Std. error</th>
<th>95% confidence interval of the difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>S.D.</td>
<td>Std. mean</td>
<td>Lower</td>
</tr>
<tr>
<td>0.02488</td>
<td>0.04255</td>
<td>0.00673</td>
<td>0.01127</td>
</tr>
</tbody>
</table>

**Abbreviations:** S.D., standard deviation; Std., standard; Sig., significance; df, degrees of freedom.
not be less than three. 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. Non-functioning distractors that are easily eliminated by students lead to easy options and causes confusion with the standard settings. 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, but they disagree with Owen and Froman. 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. 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. 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.

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

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.