Construction of Multiple Choice Questions Before and After An Educational Intervention

Karma Tenzin,[1] Thinley Dorji,[1,2] Tashi Tenzin[1,2]

[1]Faculty of Postgraduate Medicine, Khesar Gyalpo University of Medical Sciences of Bhutan, Thimphu, Bhutan, [2]Jigme Dorji Wangchuck National Referral Hospital, Thimphu, Bhutan.

ABSTRACT

Introduction: Khesar Gyalpo University of Medical Sciences of Bhutan, established in 2014, has ushered in a new era in medical education in Bhutan. Multiple Choice Questions are a common means of written assessment in medical education.

Methods: This was a quasi-experimental study conducted at the Faculty of Postgraduate Medicine, KGUMSB, Thimphu in December 2016. A total of 8 MCQs were prepared by four teaching faculties from different fields who had no prior training on construction of MCQs. It was delivered to a group of 16 randomly selected intern doctors. A 2 hours long workshop on construction of MCQs was conducted. After the workshop, the same MCQs were modified according to standard guidelines on developing MCQs and were tested in the same group of intern doctors. An analysis on the performance, difficulty factor, discrimination index and distractor analysis was done on the two sets of MCQs using Microsoft Excel and SPSS 20.0.

Results: For the pre- and post-workshop questions respectively, the pass percentage was 69.8% (11) and 81.3% (13), difficulty factor was 0.51 and 0.53, discrimination index was 0.59 and 0.47, distractor effectiveness was 83.3% and 74.9%.

Conclusions: The workshop on MCQ development apparently seemed highly valuable and effective in changing the learning and performances of medical educators in the development of MCQs.

_____

Keywords: *difficulty factor; discrimination index; faculty development; medical education.*

_____

INTRODUCTION

Multiple Choice Questions (MCQs) are a common means of written assessment in medical education. They are suitable to assess knowledge and comprehension in basic and clinical sciences.[1,2] MCQs are designed with a question and a set of responses. The correct answer is called the "key" and the others "distractors".[3] Distractors attract the students who do not know the correct answer.[4] The university or a certifying board may use MCQs as a summative assessment to certify the health professional.[2,6]

Many faculty members at the Faculty of Postgraduate Medicine, Khesar Gyalpo University of Medical Sciences of Bhutan (KGUMSB) lacked experiences in teaching and assessment techniques including development of

_____

Correspondence: Dr. Karma Tenzin, Faculty of Postgraduate Medicine, JDWNRH Campus, Thimphu, Bhutan. Email: karmatenzin9@gmail.com, Phone: +97517969450.

MCQs. The need for this was felt as academic regulation of the university mandates 50% MCQs for theoretical assessment, making it a bulk of assessment at all levels of examination at the university.

The aim of this study was to assess quality of multiple choice questions written before and after training workshop on MCQ development at the Faculty of Postgraduate Medicine, KGUMSB, Thimphu.
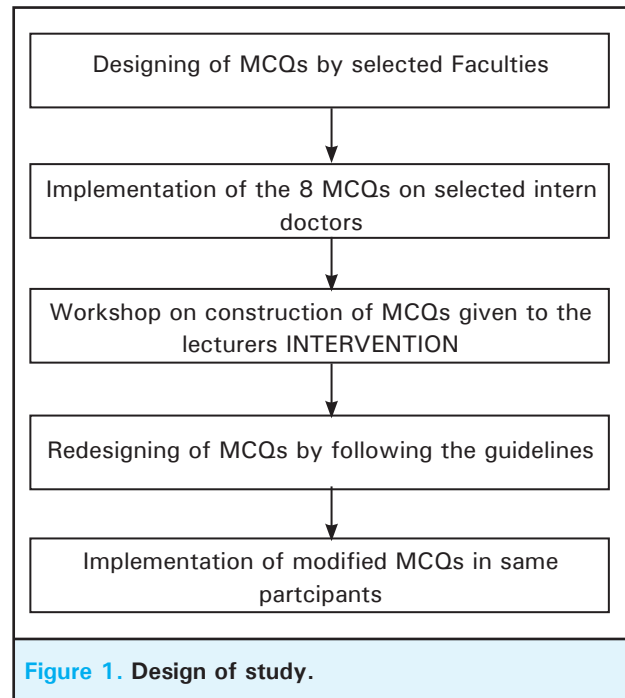
METHODS

This was a quasi-experimental study conducted at the Faculty of Postgraduate Medicine, KGUMSB, Thimphu in December 2016. The ethical approval for this study was granted by Research Ethics Board for Health, Thimphu (Ref. No. REBH/PO/2016/084 Date: 27th November, 2016). All the participants were informed regarding the purpose of the entire exercise and participation was entirely voluntary.

For this study, faculty members who had not availed training on MCQ development were included and faculties with previous exposure were excluded from the study. According to the inclusion criteria, four faculty members (single group) with no prior training, one faculty member each from Department of Physiology, Paediatrics, Pharmacology and Obstetrics and Gynaecology were included and were made to design a total of 8 MCQs. All the MCQs were best-of-four, with one key and three distractors that assessed basic knowledge and practice.

The MCQs were delivered to 16 intern medical doctors (single group) selected randomly out of 44 at the Jigme Dorji Wangchuck National Referral Hospital, Thimphu. The next day, the analysis on the face and construct validity, performance, difficulty factor, discrimination index and distractor analysis were presented as a part of the workshop on development of MCQs conducted by local experts. In addition, the workshop included exercise on development of questions with face and construct validity, development of keys and plausible distractors and analysis of student performance through difficulty factor, discrimination index, distractor effectiveness and overall reliability of the question paper through Cronbach's alpha. Guidelines set by Haladyna and the Medical Council of Canada[7, 8] were taken as standards for this workshop. Following the workshop, the faculties modified the questions with the above said guidelines. The modified MCQs were delivered again to the same group of intern doctors and results were analysed. To control the influence learning on the scores for the second set of MCQs, they were tested in the same group of intern medical doctors, the sequence of the questions was randomly mixed for each examinee,

and only 10 minutes were given to prevent recall bias.

Figure 1. Design of study.

The sampling technique adopted was convenient sampling while recruiting the faculties and simple random sampling technique while recruiting the students.

In MCQs, the analysis of items provides a measurement of quality of the questions. Data from this study was analysed in a trial version of SPSS 20.0. Difficulty factor, discrimination index and distractor analysis for each question and calculation of reliability indices was done for the pre- and post-workshop MCQs.

The difficulty factor (P) is the proportion of students who answered the question correctly. It is calculated as follows:

$$\text{Difficulty Factor} = \frac{C}{T}$$

Where, C = number selecting the correct option,
         T = total number of examinees.

The P (proportion) value statistics ranges from 0 to 1. The higher the P value, the easier the question. MCQs with a P value between 30 to 70% are considered as good and acceptable. Amongst these, items with P value between 40 to 60% are considered excellent, because the discrimination index is the maximum at this range. Items with P value less than 20% (too difficult) and more than 90% (too easy) were not accepted.[9-13]

The discrimination index (DI) is a measure of the effectiveness of an item in discriminating between high and low scorers. For this calculation, the examinees

were divided into three groups according to their scores: an upper group consisting of 4 (27%) who made the highest scores, a lower group consisting of 4 (27%) who made the lowest scores and a middle group consisting of the remaining 8 (46%).Discrimination index was estimated using the following formula:

Discrimination index $= \dfrac{P_U - P_L}{}$

Where, P_U and P_L are the proportions of the students in the upper and bottom group who got the item correct.

The range of values for the item discrimination index is −1.00 to +1.00.The higher the value of DI, the more effective the item is. When DI is 1.00, all test takers in the upper group and no test takers in the lower group answered the item correctly. Conversely, if none of the upper group but all of the lower group answered an item correctly, the DI value would be −1.00. DI <0.19 are considered poor items; DI 0.2 to 0.29 are acceptable; DI 0.3 to 0.39 are good and DI >0.4 are considered excellent. [9-13]

Non-functional distractors (NFDs) are options that are selected infrequently (<5%) by examinees and functional or effective if it is selected by >5% of examinees. Distractor efficiency (DE) is based on the number of NFDs in it and ranges from 0 to 100%. If an item contains three or two or one or nil NFDs, the DE will be 0, 33.3, 66.6, and 100% respectively. [9-13]

RESULTS

A total of 16 best-of-four multiple choice questions were studied – the pre- and post-workshop sets. The pre-workshop question paper had scores ranging from 0 to 7 with a mean of 5.1 out of 8 and a pass percentage of 69.8% (11). The post-workshop question paper had scores ranging from 1 to 7 with a mean score of 4.0 and pass percentage of 81.3% (13).

Table 1. Difficulty factor, discrimination index and distractor efficiency of each question.

	Pre-workshop MCQs			Post-workshop MCQs		
	P	DI	DE	P	DI	DE
Question 1	0.38	1.00	66.6%	0.31	0.25	66.6%
Question 2	0.81	0.75	66.6%	1.00	0	0%
Question 3	0.19	0	100%	0.25	0.25	100%
Question 4	0.19	0	100%	0.38	0.25	100%
Question 5	0.81	0.50	66.6%	0.69	0.75	66.6%
Question 6	0.31	0.75	100%	0.36	0.75	100%
Question 7	0.69	1.00	66.6%	0.56	0.75	100%
Question 8	0.69	0.75	100%	0.69	0.75	66.6%
Average	**0.51**	**0.59**	**83.3%**	**0.53**	**0.47**	**74.9%**

The difficulty factor was 0.51 and 0.53, discrimination index was 0.59 and 0.47, distractor effectiveness was 83.3% (13) and 74.9% (12) (Table 1) respectively for the pre-and post workshop MCQs.

The number of MCQs with acceptable level of difficulty (difficulty factor 30 – 70%) had increased from 4 to 6 after the workshop. The number of MCQs that were very easy, difficulty factor >70%, had decreased from 2 to 1 after the workshop. The number of MCQs that were difficult (difficulty factor <30%) had also decreased from 2 to 1 after the workshop. The questions with good discriminating power, DI ≥ 0.25, increased from 6 to 7 after the workshop. The number of questions with poor discriminating power (DI = 0 ≤ 0.20) had decreased from 2 to 1 after the workshop. On average both the pre- and post- workshops had good discriminating indices of 0.59 and 0.47 respectively. The number of questions that had 100% distractor effectiveness remained the same (four questions). However, the number of questions with distractor effectiveness of 66.6% decreased from 4 to 3 after the workshop. On average the pre- and post-workshop questions distractor effectiveness of 13 (83.3%) and 12 (74.9%).respectively.

Table 2. Level of difficulty and discriminating indices of the questions.

P score	Level of difficulty	No pre-workshop MCQs	No of post-workshop MCQs
< 30%	Difficult	2	1
30 – 60%	Ideal	2	4
60 – 70%	Acceptable	2	2
>70%	Very easy	2	1
DI	**Discriminating power**	**No pre-workshop MCQs**	**No of post-workshop MCQs**
≥ 0.25	Good	6	7
0.21 – 0.24	Acceptable	0	0
≤ 0.20	Poor	2	1
DE	**No of non-functional distractors**	**No pre-workshop MCQs**	**No of post-workshop MCQs**
0%	3	0	1
33.3%	2	0	0
66.6%	1	4	3
100%	0	4	4

DISCUSSION

MCQs have a range of advantages, such as enabling examiners to cover a large degree of content, ascertaining the correct response from the distractors. MCQs assess a large sphere of knowledge including higher cognitive skills such as application, analysis

and synthesis in Bloom's taxonomy.[5] A technical approach in development and analysis of MCQs is essential for quality assessment of the students' knowledge and practice.[7] MCQs contribute to 50% of theory paper, making one of the important and bulk of theory knowledge assessment at KGUMSB. The two sets of question papers (pre- and post-workshop) had appropriate standard of difficulty factors of 0.51 and 0.53. This measure gives a technical standing on how difficult the subject is for that particular set of students and the teachers. The papers had a discriminating power of 0.59 and 0.47, both able to identify good performers from others.

The value of the discriminating power may also be influenced by the face and construct validity. MCQs with very long stems, poorly constructed keys or distractors and vague thematic construct may fail to identify a high scorer from others. In the pre-workshop paper, at least 2 MCQs had very long stems and were poorly made with lots of irrelevant Construct (no proper objective).

Distractors were analysed to determine their usefulness in determining an examinee that has the knowledge to answer the question. The overall distractor effectiveness was 83.3% and 74.9% respectively. MCQs with higher number of NFDs (low distractor effectiveness) are easier than those with lower number of NFDs.[14] The more the number of NFDs, the more difficulty the teachers had faced in developing plausible distractors.[15] The MCQ with high number of NFDs also demonstrate the lack of alignment of answer options with the learning objective and that the answer key was too obvious even to a poor student. In terms of face and construct validity, the second set of questions contained distractors that were all plausible with identifiable thematic construct. Distractors such as "none of the above" or "always" (Absolute Quantifiers) were removed in the post-workshop questions following the Haladyna guideline.[4]

The questions that were too easy were modified to incorporate clearly defined construct themes. Two questions that were too long with vague construct (Irrelevant Construct) were improved with clearly defined construct and distractors of homogeneous content, similar length and grammatical structure. Use of negatives and double negatives in the sentences were also avoided in the second set.[4,16]

As noted by Attali and Bar-Hillel, examiners prefer to place the answer key in the middle positions (option B or C) compared to extreme positions in the ratio of 3 or 4 to 1.[17] The pre-workshop questions had their keys placed in options A, B or C and no key placed in option D. The post-workshop set had 3 keys placed in option D. It was observed that faculty development programme was of immense help in guiding the faculty members to develop effective MCQs. This study analysed only 16 MCQs, therefore it would be more accurate if a similar study with more number of MCQs and with more participants was held.

CONCLUSIONS

Faculty development program appears highly valuable and effective to change the learning and performance of medical educators in the field of development of MCQs. This workshop had marked improvement in the validity, difficulty factor, discriminating power and distractor effectiveness of the multiple choice questions. Therefore, faculty development program on development of MCQs must be given to new faculties and clinical educators joining the university.

ACKNOWLEDGEMENTS

REFERENCES

1. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9):563-7. [Full Text | PubMed]

2. Abdel-Hameed AA, Al-Faris EA, Alorainy IA. The criteria and analysis of good multiple choice questions in a health professional setting. Saudi Med J. 2005;26(10):1505-10. [Full Text | PubMed]

3. DiBattista D, Kurzawa L. Examination of the quality of multiple-choice items on classroom tests. Canadian Journal for the Scholarship of Teaching and Learning. 2011;2(2):1-22. [Full Text]

4. Haladyna TM, Downing SM. Validity of a taxonomy of multiple choice item-writing rules. ApplMeasEduc. 2002;15(3):309-34. [Full Text]

5. Palmer EJ. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? BMC Medical Education. 2007;7(49):1-7. [Full Text]

6. Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-ShaikhGK, Syed S, et al. Faculty development programs improve the quality of multiple choice questions items' writing. Scientific Reports. 2015;5(9556):1-7. [Full Text | PubMed]

7. Haladyna TM, Downing SM, Rodriguez MC. A Review

of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. ApplMeasEduc. 2002;15(3):309–34. [Full Text]

8. Medical Council of Canada. Guidelines for the development of multiple-choice questions. medical council of Canada. 2010. [Full Text]

9. 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 Ahmedabad, Gujarat. Indian J Community Med. 2014;39:17-20. [Full Text |PubMed]

10. Sim SM, 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. [Full Text | PubMed]

11. Rao C, Kishan Prasad HL, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: assessing an assessment tool in medical students. Int J EducPsychol Res. 2016;2:201-4. [Full Text]

12. Mukherjee P, Lahiri SK. Analysis of multiple choice questions (MCQs): Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. IOSR-JDMS. 2015;14(12):47-52. [Full Text]

13. Mitra N K. Nagaraja H S. Ponnudurai G. Judson J P. The levels of difficulty and discrimination indices in type A multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. IeJSME. 2009;3(1):2-7. [Full Text]

14. Abdulghani HM, Ahmad F, Ponnamperuma GG, Khalil MS, Aldrees A. The relationship between non-functioning distractors and item difficulty of multiple choice questions: A descriptive analysis. J Health Spec. 2014;2:148-51. [Full Text]

15. 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;9:40. [Full Text | PubMed]

16. Kehoe J. Writing multiple-choice test items. PARE. 1995;4:47-51. [Full Text]

17. Attali Y, Bar-Hillel M. Guess where: the position of correct answers in multiple-choice test items as a psychometric variable. J EducMeas. 2003;40(2):109-28. [Full Text]