

Replacing Non-functional Distractors to Improve the Quality of MCQs: A Quasi-experimental Study

Abdullah Aljabr¹, Saleem Shaikh², S. Karthiga Kannan³, Sami Aldhuwayhi⁴, Saikarthik Jayakumar^{5*}, Riyadh Al-Roomy⁶, Roshan Uthappa⁷ and Alhanouf Alkhujairi⁸

¹Department of Medical Education, College of Dentistry, Majmaah University, Riyadh Province, Kingdom of Saudi Arabia

^{2,3}Department of Maxillofacial Surgery and Diagnostic Sciences, College of Dentistry, Majmaah University, Riyadh Province, Saudi Arabia

⁴Department of Prosthodontics, College of Dentistry, Majmaah University, Riyadh Province, Saudi Arabia

⁵Department of Basic Medical Sciences, College of Dentistry, Majmaah University, Riyadh Province, Saudi Arabia

^{6,7}Department of Restorative Dental Sciences, College of Dentistry, Majmaah University, Riyadh Province, Saudi Arabia

⁸Department of Oral and Maxillofacial Surgery, College of Dentistry, Majmaah University, Riyadh Province, Saudi Arabia

KEYWORDS Difficulty Index. MCQ. Non-functional Distractor. Post Exam Analysis

ABSTRACT The quality of multiple-choice questions (MCQs) can be determined by the discrimination index (DsI), difficulty index (DI), and distractor efficiency (DE) in post exam analysis. Selection of distractors is one of the most neglected aspects of MCQ preparation. The present study evaluated the effect on DsI by replacing non-functional distractor (NFD) with more plausible alternatives using a quasi-experimental study model. The DsI and NFD values were obtained from the MCQ examinations that were conducted during the year 2017 from 37 courses with each containing 50 MCQ's in the College of Dentistry, Majmaah University. The post exam analysis results were informed to the course directors and the faculty were given orientation program about replacing the NFD's with more plausible options. In the following year (2018), sixty percent of the previous year MCQ's were used with modified NFD's in those questions (1100 MCQ's). The average DsI and NFD were significantly improved from the year 2017 to 2018 ($P < 0.000$) with significant negative correlation between them ($P < 0.000$). This study shows that the psychometric properties of MCQs can be improved by creating better distractors. The researchers recommend providing faculty with more training on how to use distractors effectively in MCQ writing.

INTRODUCTION

Multiple choice questions (MCQs) or single best answers (SBAs) are among the most widely used assessment tools and often account for a major portion of course grades (DiBattista and Kurzawa 2011; Mahjabeen et al. 2017). MCQs are very popular among university faculty because they can be standardised easily, they are objective, valid and reliable, and evaluator bias is eliminated, making them very useful in high stakes

tests, such as professional qualification exams (Considine et al. 2005).

To enhance the quality of exams, MCQ test items are analysed using information from students' responses to evaluate the quality of test items (Collins 2006; Considine et al. 2005; Shaikh et al. 2020). Parameters like the difficulty index (DI), discriminatory index (DsI), distractor efficiency (DE), coefficient of reliability, etc., help to identify questions requiring amendment before item banking or that need to be deleted (Chavda et al. 2015; Zubairi and Kassim 2006). An important factor determining the quality of an MCQ is its DsI, which tells one the extent to which a more knowledgeable student is likely to answer the question correctly than a less knowledgeable student. The value of DsI should be positive, which means that the proportion of students who obtained higher grades also performed better on this particular question.

*Address for correspondence:

Dr Saikarthik Jayakumar
Department of Basic Medical Sciences, College of Dentistry
Majmaah University
Al Zulfi, Riyadh Province
Saudi Arabia -11932
Fax: 064227184
Mobile: 966531272304
E-mail: s.jaya@mu.edu.sa

An item's DsI, in turn, depends heavily on the quality of distractors. Lowe (1991) noted, "The mark of a good question is often the quality of the distractors," and this reflects the importance of these incorrect options for discrimination between examinees. A good distractor should seem plausible to less knowledgeable students but should not mislead more knowledgeable students away from the correct answer. A distractor is said to be effective when at least some examinees select it, and if it is not selected, then it is non-functional and does not contribute to the discriminatory power of the MCQ. Many authors agree that at least five percent of examinees must select a distractor for it to be considered functional (Tarrant et al. 2009). Analysis of distractors can serve as a valuable guide to the quality of MCQs, as the distractors should create reasonable doubt about which option is the correct answer. Even though some authors have mentioned that the quality of distractors is the most important part of item analysis, this remains one of the most neglected aspects of MCQ writing (Haladyna and Downing 1989; Shenoy et al. 2016). Most faculty members spend a lot of time and energy on the construction of stem and lead in parts of an MCQ but use distractors only as fillers to complete the questions. Additionally, creating options with equal plausibility in the health sciences is a difficult task. Some studies have investigated the role of NFDs on discrimination and difficulty indexes by reducing the number of options from five to three (Deepak et al. 2015; Elfaki et al. 2015). However, the results have been inconsistent and many academics feel that more options help them to cover more content, and additionally, three-option MCQs increase weaker students' chances of getting correct answers by guessing (Rogers and Harley 1999; Trevisan et al. 1991). Similarly, replacing and correcting the faulty distractor could positively improve the discrimination indices (Ali and Ruit 2015).

Objectives

As limited studies exist in the literature on the qualitative assessment of MCQ, the present study was done to evaluate the effect of replacing NFDs with more reasonable or plausible alternatives.

MATERIAL AND METHODS

This study was conducted by the Department of Dental Education, College of Dentistry at Majmaah University in Al Majmaah City, Kingdom of Saudi Arabia. Approval for the study was obtained by the institutional review board of the university. Completed examinations consisting of 50 MCQs each were obtained from 37 courses across all levels of the program in 2017, and item analysis was performed. There were a total of 1850 MCQ's across all courses (37 * 50) and the DsI and NFD's were recorded. The course instructors of all 37 subjects were then made aware of the results of the item analysis and were suggested to focus on replacing the NFDs with more plausible options on the exam in the following year. For the 2018 exam, the course directors of the same 37 courses were instructed to retain at least sixty percent (1100 MCQ's) of the questions from the 2017 examination papers after changing the NFDs. The following MCQ is taken as an example to explain this.

'Which of the following cells are the most prominent cells seen in acute inflammation?'

- a. Neutrophils
- b. Lymphocytes
- c. Epithelial cells
- d. Blood vessels'

In this question, the researchers found that options C and D were not inflammatory related cells and hence the faculty were given feedback to change distractors. On post exam analysis this question had two NFDs and DsI of 15 and DE of 33.3 percent. In the year 2018, the same question was repeated with modified distractors as below.

'Which of the following cells are the most prominent cells seen in acute inflammation?'

- a. Neutrophils
- b. Monocytes
- c. Lymphocytes
- d. Eosinophil'

Post exam analysis revealed that there was 0 NFD with DsI at 82 and DE of one hundred percent.

For the purpose of the study, only those 1100 MCQ's, which were common between 2017 and 2018, were evaluated. The content validity of the questions was verified by the departmental council and approved for both examinations.

The DsI measures how examinees with high scores perform on a particular MCQ item compared to examinees with low scores on the same

item. The DsI scores ranged between -100 and 100. A positive value indicates that the student who performed well on the total test also answered the item correctly. Conversely, a negative value suggests that the student, who scored high on the total test, answered the item incorrectly. A high (or positive) score indicates good discrimination, whereas a negative score indicates an unacceptable question. In the present study, an item with a DsI score of more than 75 was graded as good, a score between 35 and 75 was graded as moderate, while items with scores of less than 35 were graded as having poor discriminatory capacity. The items with negative DsI scores were noted separately.

Distractors that were not chosen by at least five percent of the examinees were considered to be an NFD (Tarrant et al. 2009). The number of NFDs in each question was noted and the questions were categorised as containing 0, 1, 2 or 3 NFDs. An item with 0, 1, 2 and 3 NFD's has a distractor efficiency of 100.0 percent, 66.66 percent, 33.33 percent and 0 percent, respectively (D' Sa and Visbal-Dionaldo 2017). It has been reported that items having 0 or 1 NFD have good discriminating ability than items with 2 or 3 NFDs,

Hence the MCQs in this study were grouped into 2 sets, one having MCQs with 0 and 1 NFD while the other had MCQs with 2 or 3 NFDs. The same 37 courses that were selected for the 2017 examinations were analysed for DI, DsI, and proportion of NFDs. The DsI and NFDs along with results of examination were provided by a Scantron OCR scanner.

Statistical Analysis

The DsI and NFD across 2017 and 2018 for all 37 subjects were analysed using descriptive statistics and Chi-square test. The effect sizes were also calculated. All statistical analysis was performed using SPSS (USA, Version 26). P value < 0.05 is considered significant.

RESULTS

It was noted that significant difference existed in the distribution of MCQs with negative, less than 35.0, 35.0 to 75.0 and more than 75.0 percent of DsI between the year 2017 and 2018 (Table 1; $\chi^2=269.01$, $P<0.001$). Significant differences were noted in the distribution of MCQs

Table 1: Distribution of Discrimination Index (DsI)

Discrimination	Actual 2017	Percentage 2017	Actual 2018	Percentage 2018	Chi-square statistic	Significance	Effect size (r)
Negative	37	3.36	28	2.52	269.01	.00001	0.5 large
Less than 35	772	69.57	402	36.22			
35 - 75	257	23.11	537	48.37			
More than 75	44	3.96	143	12.88			
Total	1110	100	1110	100			

Note: Chi-square test shows significant difference in the distribution of negative, less than 35, from 35 to 75, more than 75 (percent) DsI between the year 2017 and 2018. The effect size was 0.5, indicative of large effect size. The result is significant at $p < .05$. It can be noted that after the feedback to the course directors, the discrimination index improved in the year 2018.

Table 2: Distribution of Non-functional Distractors (NFDs)

No. of NFDs	Actual 2017	Percentage 2017	Actual 2018	Percentage 2018	Chi-square statistic	Significance	Effect size (r)
0	146	13.2	205	18.48	57.59	0.00001	0.28 medium
1	299	26.9	376	33.87			
2	347	31.26	350	31.53			
3	318	28.64	179	16.12			
Total	1110	100	1110	100			

Note: Chi-square test shows significant difference in the distribution of MCQ's with 0, 1, 2 and 3 NFD's. The effect size was moderate. The result is significant at $p < .05$. It can be noticed that the percentage of NFD has significantly reduced in the year 2018.

with 0, 1, 2, 3 NFD's when compared between two years ($\chi^2=57.59$, $P<0.001$) and with moderate effect size ($r=0.28$) (Table 2). Significant difference was also noted in NFD's when they were grouped into 0, 1 and 2, 3 between 2017 and 2018 ($\chi^2=33.52$, $P<0.001$). However, the effect size was low ($r=0.17$) (Table 3) while Table 4 shows the correlation between DsI and NFDs. The results showed a strong negative correlation ($r=-0.821$) between DsI and NFDs, showing that higher the average number of NFD's, the lower is the DsI (Table 4).

DISCUSSION

Assessment, whether formative or summative, is an important aspect of education, and the quality of assessment to a large extent also determines the quality of students' knowledge and skills upon completion of a course. One of the major indicators of a good assessment is its ability to discriminate between good students and poor students (Kilgour and Tayyaba 2016). This study evaluated the number of NFDs in each item for the exams conducted in 2017 and found that only 13.2 percent of the questions had no NFDs (discrimination efficiency (DE): 100 percent), 26.9 percent had one NFD, 31.26 percent had two NFDs, and 28.64 percent of the items had three NFDs (DE=0). These figures were

either similar or inferior to most of the studies reporting the prevalence of NFDs (Haladyna and Downing 1993; Tarrant et al. 2009; Vyas and Supe 2008). There is substantial evidence to suggest that most MCQs used in examinations contain a high proportion of NFDs (Baghaei and Amrahi 2011; Cizek and O'Day 1994; Rogausch et al. 2010; Tarrant et al. 2009). Al Rukban (2006) opined that the process of generating distractors for MCQs is complex and time consuming, and this often results in the selection of unconvincing distractors. Some researchers have even suggested that in the health sciences, there is a natural limit to the number of plausible distractors for any given topic, and that this limit is generally less than four (Haladyna and Downing 2002). Based on this argument, many studies have removed the poor-quality distractors using a reduced option format (Kolstad et al. 1985; Landrum et al. 1993; Rahma et al. 2017; Schneid et al. 2014; Shizuka et al. 2006). The early studies tested the effect of removing the poor options. In Sidick et al. (1994), the authors, with the help of experts, removed poorly functioning options and created three-option MCQs from five-option MCQs. They found that the removal of the non-functioning options did not influence the psychometrics. The authors, therefore, concluded that three options are adequate. Rodriguez (2005) suggested that three options are optimal

Table 3: Distribution of MCQs with 0, 1 and 2, 3 NFDs

No. of NFDs	Actual 2017	Percentage 2017	Actual 2018	Percentage 2018	Chi-square statistic	Significance	Effect size (r)
0/1	445	40.09	581	52.34	33.52	.00001	0.17
2/3	665	59.9	529	47.64			small
Total	1110	100	1110	100			

Note: It can be noted here that there was a significant difference in the distribution of MCQs between these two categories with smaller effect sizes. The percentage of MCQs with 2 or 3 distractors is significantly less than the percentage of MCQs with 0 or 1 distractor.

Table 4: Correlation between Average discrimination and average NFD

Descriptive Statistics	Correlations				
	Mean	Std. Deviation	N	Pearson Correlation	Sig. (2-tailed)
Average discrimination	33.6527	8.408	37	-0.821	.000
Average NFD	1.5343	0.38616			

Note: It can be noted that strong significant negative correlation was present between average DsI and average NFDs. The result is significant at $p < .05$.

in any MCQ, and they can perform as well as four options with less time required by faculty to make tests. Haladyna and Downing (1988) suggested in their study that two distracters play a more effective role in MCQs than four or five options because distracters beyond three options are mostly “fillers”. Some studies showed improvement in the quality of questions, whereas many did not show any major difference in the item parameters (Cizek et al. 1998; Owen and Froman 1987).

Despite the argument that it is difficult to make four-option MCQs and that a reduced format should be used, major testing organisations, MCQ writing guidebooks, institutional policies, and MCQ banks mostly use four- or five-option MCQs (Ercikan and Julian 2002; Masters et al. 2001). This could be because the teachers find more options help in good content coverage and using three-option MCQs increases the chances of correct guessing by weaker students. Additionally, assessments are conducted according to the established rules of the educational institution and teachers generally have very little role in deciding the format. The item writers often find it difficult to strike a balance between efforts to generate several functioning options and while reducing the number of total options (Swanson et al. 2008; Vyas and Supe 2008).

In this study, the researchers did not delete the NFDs. Rather, they chose to replace them with more plausible distractors with the help of subject experts. When these items were tested again in the next exam, it was noticed that the questions with no NFDs or one NFD had increased to 18.5 percent and 33.87 percent respectively, and the items with three NFDs had decreased to 16.12 percent. These psychometric properties of the items had significantly improved from the previous year and were also better than many in studies mentioned earlier. The study showed that in 2017, 40.09 percent of the questions had fewer than two NFDs and 59.9 had more than two NFDs. Following the intervention by the medical education department in 2018, the study found that more than half of the MCQs (52.34%) had fewer than two NFDs, which is very desirable, and only 47.34 percent had more than two NFDs. These results may require further improvement when compared to other studies, but the researchers could show

a statistically significant improvement by replacing the distractors rather than by deleting them (Abdulghani et al. 2015; Tenzin et al. 2017).

Although this study used a novel approach to improve the quality of MCQs by replacing NFDs, one of the limitations of the study was that the researchers used only one criterion to define NFD and did not consider the amount of time spent on each distractor or the weight, which would have defined the non-functionality of a distractor more effectively.

CONCLUSION

The results of the study confirm that the psychometric properties of an MCQ can be improved by creating better distractors, and that reducing the number of options may not always be required. Therefore, the researchers suggest that the quality of distractors is more relevant than their quantity. One of the reasons for the high number of NFDs seen in many previous studies is the lack of faculty training and professional developmental programs, which prevents the faculty from being equipped with MCQ writing skills. The results of this study would enable faculty members to be aware of flaws in the MCQ and address them in a scientific manner.

RECOMMENDATIONS

The result of this study warrants the educational institutions implement faculty training in MCQ assessments with emphasis on the fact that distractors are an important component of MCQs and not just fillers, as this aspect is more often neglected.

ACKNOWLEDGEMENTS

The researchers would like to thank the Dean-ship of Scientific Research, Majmaah University for the support to this academic project. They would also like to thank Editage (www.editage.com) for the English language editing.

REFERENCES

- Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK, Syed S et al. 2015. Faculty development programs improve the quality of multiple choice questions items' writing. *Scientific Reports*, 5: 9556. DOI: 10.1038/srep09556

- Al-Rukban MO 2006. Guidelines for the construction of multiple choice questions tests. *Journal of Family Community Medicine*, 13(3): 125–133.
- Ali SH, Ruit KG 2015. The impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. *Perspectives on Medical Education*, 4(5): 244–251.
- Baghaei P, Amrahi N 2011. The effects of the number of options on the psychometric characteristics of multiple choice items. *Psychological Test and Assessment Modelling*, 53(2): 192–211. DOI: 10.1007/s10459-015-9652-7
- Chavda P, Misra S, Duttaray B 2015. Item analysis of MCQs. *South East Asian Journal of Medical Education*, 9(1): 66–68.
- Cizek GJ, O'Day DM 1994. Further investigation of nonfunctioning options in multiple-choice test items. *Educational and Psychological Measurement*, 54(4): 861–872. DOI: 10.1177/0013164494054004002
- Cizek GJ, Robinson KL, O'Day DM 1998. Non-functioning options: A closer look. *Educational and Psychological Measurement*, 58: 605–611.
- Collins J 2006. Education techniques for lifelong learning: Writing multiple-choice questions for continuing medical education activities and self-assessment modules. *Radiographics: A Review Publication of the Radiological Society of North America*, 26(2): 543–551. DOI: 10.1148/rg.262055145
- Considine J, Botti M, Thomas S 2005. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian Journal of the Royal College of Nursing Australia*, 12(1): 19–24. DOI: 10.1016/S1322-7696 (08)60478-3
- Deepak KK, Al-Umran KU, Al-Sheikh MH, Dkoli BV, Al-Rubaish A 2015. Psychometrics of multiple choice questions with non-functioning distractors: Implications to medical education. *Indian Journal of Physiology and Pharmacology*, 59(4): 428–435.
- DiBattista D, Kurzawa L 2011. Examination of the quality of multiple-choice items on classroom tests. *The Canadian Journal for the Scholarship of Teaching and Learning*, 2(2): 1–23. DOI: 10.5206/cjsotl-rcaeca.2011.2.4
- D'Sa JL, Visbal-Dionaldo ML 2017. Analysis of Multiple Choice Questions: Item Difficulty, Discrimination Index and Distractor Efficiency. *International Journal of Nursing Education*, 9(3): 109–114. doi:10.5958/0974-9357.2017.00079.4.
- Elfaki OA, Bahamdan KA, Al-Humayed S 2015. 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*, 10(4): 123–127.
- Ercikan K, Julian M 2002. Classification accuracy of assigning student performance to proficiency levels: Guidelines for assessment design. *Applied Measurement in Education*, 15(3): 269–294. DOI: 10.1207/S15324818AME1503_3
- Haladyna TM, Downing SM 1988. Functional Distractors: Implications for Test-item Writing and Test Design. *Paper presented at Annual Meeting of the American Educational Research Association*, New Orleans, 5-9 April.
- Haladyna TM, Downing SM 1989. Validity of taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1): 51–78. DOI: 10.1207/s15324818ame0201_4
- Haladyna TM, Downing SM 1993. How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement*, 53: 999–1010.
- Haladyna TM, Downing SM, Rodriguez MC 2002. A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3): 309–333. DOI: 10.1207/S15324818AME1503_5
- Kheyami D, Jaradat A, Al-Shibani T, Ali FA 2018. Item analysis of multiple choice questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos Univ Med J*, 18(1): e68-e74. DOI: 10.18295/squmj.2018.18.01.011.
- Kolstad RK, Briggs LD, Kolstad RA 1985. Multiple-choice classroom achievement tests: Performance on items with five vs three choices. *College Student Journal*, 19(4): 427–431.
- Kilgour JM, Tayyaba S 2016. Investigations into the optimal number of distractors in single-best answer exams. *Advances in Health Science Education*, 21(3): 571–585. DOI: 10.1007/s10459-015-9652-7
- Landrum RE, Cashin JR, Theis KS 1993. More evidence in favor of three-option multiple-choice tests. *Educational and Psychological Measurement*, 53(3): 771–778.
- Lowe D 1991. Set a multiple choice question (MCQ) examination. *BMJ*, 302(6779) 780–782.
- Mahjabeen W, Alam S, Hassan U, Zafar T, Butt R, Konain S et al. 2017. Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Annals of PIMS*, 13(4): 310–315.
- Masters JC, Hulsmeyer BS, Pike ME, Leichty K, Miller MT, Verst AL 2001. Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. *Journal of Nursing Education*, 40(1): 25–32.
- Owen SV, Froman RD 1987. What's wrong with three-option multiple-choice items? *Educational and Psychological Measurement*, 47: 513–522. DOI: 10.1177/0013164487472027
- Rahma NAA, Shamad MMA, Idris MEA, Elfaki OA, Elfakey WEM, Salih KMA 2017. Comparison in the quality of distractors in three and four options type of multiple-choice questions. *Advances in Medical Education and Practice*, 8: 287–291. DOI: 10.2147/AMEP.S128318
- Rodriguez MC 2005. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2): 3–13. DOI: 10.1111/j.1745-3992.2005.00006.x
- Rogausch A, Hofer R, Krebs R 2010. Rarely selected distractors in high stakes medical multiple choice examinations and their recognition by item authors: A simulation and survey. *BMC Medical Education*, 10: 85. DOI: 10.1186/1472-6920-10-85

- Rogers WT, Harley D 1999. An empirical comparison of three- and four-choice items and tests: Susceptibility to testwiseness and internal consistency reliability. *Education and Psychological Measurement*, 59(2): 234–247. DOI: 10.1177/001316499219 69820
- Schneid SD, Armour C, Park YS, Yudkowsky R, Bordage G 2014. Reducing the number of options on multiple-choice questions: Response time, psychometrics and standard setting. *Medical Education*, 48(10): 1020–1027. DOI: 10.1111/medu.12525
- Shaikh S, Kannan SK, Naqvi ZA, Pasha Z, Ahamad M 2020. The role of faculty development in improving the quality of multiple choice questions in dental education. *J Dent Educ*, 84: 316–322. doi:10.21815/JDE.019.189
- Shenoy PJ, Sayeli V, Rao RR 2016. Item-analysis of multiple choice questions: A pilot attempt to analyze formative assessment in pharmacology. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, 7: 1683–1687.
- Shizuka T, Takeuchi O, Yashima T, Yoshizawa K 2006. A comparison of three- and four-option English tests for university entrance selection purposes in Japan. *Language Testing*, 23(1): 35–57. DOI: 10.1191/026 553 2206lt319oa
- Sidick JT, Barrett GV, Doverspike D 1994. Three alternative multiple choice tests: An attractive option. *Personnel Psychology*, 47: 829–835. DOI: 10.1111/j.1744-6570.1994.tb01579.x
- Swanson DB, Holtzman KZ, Allbee K 2008. Measurement characteristics of content-parallel single-best-answer and extended-matching questions in relation to number and source of options. *Academic Medicine*, 83(10): S1–S4. DOI: 10.1097/ACM. 0b013e 318183e5bb
- Tarrant M, Ware J, Mohammed AM 2009. An assessment of functioning and non-functioning distracters in multiple-choice questions: A descriptive analysis. *BMC Medical Education*, 9(40): 1–8. DOI: 10.1186/1472-6920-9-40
- Tenzin K, Dorji T, Tenzin T 2017. Construction of multiple choice questions before and after an educational intervention. *Journal of the Nepal Medical Association*, 56(205): 112–116. DOI: 10.31729/jnma.2976
- Trevisan MS, Sax G, Michael WB 1991. The effects of the number of options per item and student ability on test validity and reliability. *Educational and Psychological Measurement*, 51(4): 829–837. DOI: 10.1177/001316449105100404
- Vyas R, Supe A 2008. Multiple choice questions: A literature review on the optimal number of options. *The National Medical Journal of India*, 21(3): 130–133.
- Zubairi AM, Kassim NLA 2006. Classical and Rasch analyses of dichotomously scored reading comprehension test items. *Malaysian Journal of ELT Research*, 2(1): 1–20.

Paper received for publication in April, 2021
Paper accepted for publication in June, 2021