DOI: https://doi.org/10.54393/pjhs.v4i04.700

PAKISTAN JOURNAL OF HEALTH SCIENCES

https://thejas.com.pk/index.php/pjhs

Volume 4, Issue 4 (April 2023)

Original Article

Improving the Quality of MCQs by Enhancing Cognitive Level and using Psychometric Analysis

Muhammad Faisal Rahim[1*], Shireen Qassim Bham[2], Shahjabeen Khan[3], Tahir Ansari[1] and Masood Ahmed[4]

[1]Department of Medical Education, Fazaia Ruth Pfau Medical College, AIR University, Karachi, Pakistan
[2]Department of Paediatrics, Fazaia Ruth Pfau Medical College, AIR University, Karachi, Pakistan
[3]Department of Physiology, Fazaia Ruth Pfau Medical College, AIR University, Karachi, Pakistan
[4]Department of Anatomy, Fazaia Ruth Pfau Medical College, AIR University, Karachi, Pakistan

ARTICLE INFO

Key Words:
Item Analysis, Cognition, Discrimination Index, Difficulty Index, Distractor

How to Cite:
Rahim, M. F. ., Qassim Bham, S. ., Khan, S., Ansari, T. ., & Ahmed, M. . (2023). Improving the Quality of MCQs by Enhancing Cognitive Level and using Psychometric Analysis: Improving the quality of MCQs by enhancing cognitive level. Pakistan Journal of Health Sciences, 4(04), 115–121. https://doi.org/10.54393/pjhs.v4i04.700

***Corresponding Author:**
Muhammad Faisal Rahim
Department of Medical Education, Fazaia Ruth Pfau Medical College, AIR University, Karachi, Pakistan
mfaisalrahim@gmail.com

Received Date: 10th April, 2023
Acceptance Date: 28th April, 2023
Published Date: 30th April, 2023

ABSTRACT

Faculty development programs are an essential part of any reputable institution. Faculty training through various conferences and workshops will help develop competencies for high quality assessment. **Objective:** To evaluate the effectiveness of ongoing faculty training initiatives aimed at improving the standard of MCQ development. **Methods:** Faculty members of FRPMC were engaged in this observational, quasi-experimental study. Convenient sampling was done. Three different questionnaires were provided to the participants. The participants were tasked with accurately identifying flaws, cognitive levels in the pre- and post-test questionnaire items, as well as post hoc analysis to discard or modify the Items. **Results:** Items with multiple flaws were assessed with mean value in pre-test score was 1.349 compared to post-test score 3.442 which were statistically significant (p-value <0.05). The pre- and post-test questionnaire to correctly identify the cognitive levels showed pre-test 77.5% compared to 87.5% participants identified correct cognitive level in post-test. In post hoc analysis, participants were able to identify 5 questions out of 7 which need to be discarded and 3 questions out of 5 which needed improvement according to the key given to them. **Conclusions:** Our study revealed that workshop helped faculty identify item flaws with significant improvement in pre- and post-test scores. The cognitive capabilities of faculty were improved in all three levels of cognition, with greatest improvements shown in higher complexity questions (C3). There was a noticeable improvement in knowledge and the participant's capacity to identify and correct errors.

INTRODUCTION

Assessment is a very dominant force which drives learning and acts as a prime motivator for students' performance for all the teaching and learning that takes place in the medical curriculum. The main purpose of assessment is to provide feedback on the process of learning [1]. Furthermore it also helps in making a choice of assessment methods which would best suit learning. It is important, whichever assessment method is used, it must be valid, reliable, fair, feasible and objective [2]. A wide range of assessment methods are available in medical education which include short essay questions, short answer questions, multiple choice questions, checklists, objective structured practical examination (OSPE), objective structured clinical examination (OSCE), direct observation of procedural skills (DOPS), objective structured assessment of technical skills (OSTATS), short case, long case, extended matching items, portfolios, log book, simulators, peer assessment and standardized patients (SP) [3]. Multiple-choice questions (MCQs) are the most frequently used method for written assessment. It measures different cognitive levels ranging

DOI: https://doi.org/10.54393/pjhs.v4i04.700

from recall, comprehension to analysis and can be designed to measure application of knowledge according to bloom's taxonomy. MCQs have better validity, reliability and objectivity when compared to other written test assessment methods such as True/False questions, SAQs, SEQs, long essay questions [4]. MCQs development is a time consuming exercise. A typical approach of developing MCQs include selection of the topic, the development of a stem or a case, lead-in question, list of distractors or options and a key. It is imperative to review each MCQs to remove flaws and errors for further improvement in the quality of the MCQs [5]. It is not possible for anyone to develop a good quality MCQ without a thorough peer review. For this purpose a review panel of experts can be established who conscientiously review and remove potential flaws in the items to improve the quality of MCQ based written assessment [6]. The first element affecting the quality of MCQs is the absence of a higher cognitive level. Most of the time, C1-level questions or recall questions are asked in the exams [7]. These questions assess only surface knowledge and check students' capability of memorization. They are unable to assess students' deep knowledge and problem solving skills as they did not use clinical vignette or problem based questions in them. According to Pickford and Newcomb, C1 level questions in MCQs should only be used to assess factual knowledge [8]. In the modern curriculum it should be discouraged as much as possible. Higher cognition in MCQs can be achieved through using interpretation, comprehension, analysis, synthesis, correlations, clinical decision making and problem solving questions according to Modified bloom's taxonomy [9]. Higher cognitive questions help in building more practical, clinically relevant questions in the exams. The higher cognitive questions are often referred to as C2 and C3 questions. The paucity of C3 problem-based questions is a result of the fact that most examiners found it challenging to create these types of questions, which need a lot of faculty time and effort. In addition, discussions and consultations with an interdisciplinary team were also required to develop these types of questions. As most of the exams in our current setting consist of C1 questions, as students prefer surface or factual knowledge, recall questions in the exams were easier to answer and produce better results in exams for the students and possess less burden on faculty [10]. Realizing that these questions only check recall and surface knowledge, it is highly recommended to construct questions that need higher cognitive abilities in order to assess students' problem-solving and deep learning strategies [11]. The second most important factor which compromises the quality of MCQs is the presence of unstable data which is the presence of flaws and errors in the MCQs [12]. These errors or flaws are usually, presence of spelling or grammatical mistakes, use of absolute terms, use of vague terms, implausible distractors, absence of lead in statement, use of extra details or presence of long statement in option, use of all of the above or none of the above in the statement option, negative stem, lack of homogeneity in the options, use of jargons, presence of cues leading to correct answers, word repeats are some of the major flaws present in MCQs [13]. Third aspect which would be considered is to evaluate previously held exams post-hoc analysis reports. Post hoc analysis of items is done through psychometric analysis. This Psychometric analysis of Items or MCQs in a post-hoc analysis report identifies discrepancies of test items. This type of analysis helps medical educationists and faculty to improve the validity and reliability of the exam [14]. Psychometrics analysis of Items is basically composed of difficulty index, discrimination index, and distractor analysis. Normally items too difficult or items too easy are discarded from paper. In other words items which have a very low difficulty index or with a very high difficulty index are excluded from the exam paper. Items which were unable to discriminate between a good performer and poor performer were also discarded or marked as revision required and is called discrimination index. Distractor analysis analyzes the presence of noisy distractors and silent distractors. The options in the MCQs which distract students more than a key or correct option are not required, similarly options which are silent or do not participate to distract examinee were also eliminated [15]. Furthermore, to check the internal consistency of the MCQs paper, Cronbach's alpha values are calculated. If reliability coefficient does come under acceptable range, those items need further improvement or be discarded from the exam. All of these psychometric parameters of item analysis mentioned above improves the reliability and validity of the exam [16].

The mission of Fazaia Ruth Pfau Medical College (FRPMC), a newly established medical college is to produce competent physicians through excellence in medical education and research and to produce future doctors who are community oriented and socially accountable. Hence, the strategies of assessment methods adopted in the curriculum should also depict the same institutional philosophy. In order to achieve this, each assessment instrument should be designed which allows students to be self-directed learners, lifelong learners, problem solvers, practical in their approach. This is accomplished by modifying and raising the cognitive level of our assessment methodologies. Without creating cohesive faculty development programmes for evaluation methodologies, this would be a very difficult task. At FRPMC, our medical educationists regularly train faculty in house through

DOI: https://doi.org/10.54393/pjhs.v4i04.700

certificate courses, Continuous medical education meetings and seminars. In addition to this faculty trained through various faculty training programs outside college through conferences and workshops to develop competencies for high quality assessment. The aim of our study was to see the impact of ongoing training programs of faculty to improve the quality of MCQs development. Faculty would be able to achieve the following objectives through training (i) develop higher cognitive level in MCQs development (ii) identification of various items flaws and errors in MCQs (iii) eliminate the items on the basis of Item analysis and distractor analysis (iv) retain only items with good reliability coefficient.

METHODS

The study design was observational, quasi-experimental involving faculty of Fazaia Ruth Pfau Medical college, 100 faculty members took part in this study including faculty from basic and clinical health sciences from junior to senior level. The participants were the faculty members who have not attended any workshop on MCQ development previously or those faculty members who have attended workshop but still lacking training in MCQ development. Observers were from the faculty of medical education experienced in conducting workshops in medical education. The study period was from January 2021 till April 2023. Convenient sampling was done and data was analyzed using pre- and post- testing through paired t-test in SPSS version 20.0 with a quasi-experimental pre- and post-test study design. The faculty were given pre-test questionnaires before the start of workshop and post-test questionnaires were given to participant at the end of study. Three different objectives were studied using three different questionnaires, to check capacity of faculty members to correctly identify various attributes which were involved in improving the quality of MCQs. The first questionnaire was given to 50 participants who were assigned to correctly identify flaws, errors in the items. Pre-test questionnaire portion was given before the training program while the post-test portion of the questionnaire was given after the training. There were 25 items (MCQs) each containing 3-6 item flaws. Faculty are given a task to correctly identify flaws and errors from each item. See Tables 1 for list of Flaws and Errors in MCQs from where the questionnaire was derived. After pre-test, faculty were trained in faculty training session where they were told about all possible flaws and errors and after the session post-test portion of questionnaire was distributed. This post-test questionnaire was the same as the pre-test questionnaire. Both pre-post results were compared in SPSS as paired t-test.

Table 1: List of Flaws and Error in MCQs provided to Faculty with the questionnaire

S. No	List of Flaws and Errors in MCQs
1.	Lack of leading Stem
2.	Lack of Clinical vignette
3.	Grammatical mistakes
4.	Ambiguous or unclear language
5.	Use of All of Above
6.	Use of None of Above
7.	Use of Vague terms
8.	Negatively framed question
9.	True and False or Fill in the Blank type questions
10.	Use of Abbreviations and Jargons
11.	Use of Absolute terms
12.	Problem in stem is not related to options given
13.	Presence of long option as cue
14.	Repeat of sentence in stem and option providing cue
15.	Lack of homogeneity in the options
16.	Very Complex and complicated stem or question

The second pre- and post-test questionnaire comprised 12 MCQs in which another 50 faculty members were assigned to correctly identify the cognitive levels (C1, C2, C3). Like before, pre- and post-test results of both questionnaires were compared in SPSS using paired t-tests. Third questionnaire was given in the form of a report of the 2nd year MBBS Nutrition module consisting of post hoc item analysis. This questionnaire was given to all 100 participants. This report was converted into a questionnaire in which details of item analysis for each MCQ or item were present (Means Difficulty index, Discrimination Index and Option analysis were already calculated for each item). Faculty were assessed to make decisions regarding which questions (MCQs) to keep, reject, or mark for revision based on the data presented to them.

RESULTS

Results of the Table 2 show comparison of mean scores of pre-test and post-test of the participants who were trained in the MCQ development workshop. The test consisted of 25 questions and each question had multiple flaws. The test score was based on the number of flaws identified by the participants and scores were calculated according to the key compiled by the trainers. There is a significant difference in the scores of pre-test when compared to post-test after the training workshop. The difference was comparatively less in five items (Item No: 17, 18, 19, 23, 25) but still statistically significant (<0.05). Therefore it can be inferred that participants' ability to identify flaws and errors in MCQs improved after attending the training program.

DOI: https://doi.org/10.54393/pjhs.v4i04.700

Table 2: Frequencies of Detection of Flaws & Errors in the items. (n=50)

Q. #	No. of Flaws	Mean of Pre-test Score	Mean of Post-test Score	Mean Difference	p-Value
1	4	0.84	3.40	2.56	0.000
2	5	0.72	3.40	2.68	0.000
3	6	1.76	4.40	2.64	0.000
4	6	2.20	3.72	1.52	0.000
5	5	1.00	3.52	2.52	0.000
6	4	1.24	3.76	2.52	0.000
7	5	0.56	3.48	2.92	0.000
8	5	0.44	3.60	3.16	0.000
9	5	0.80	3.76	2.96	0.000
10	3	1.24	2.84	1.60	0.000
11	8	1.84	3.88	2.04	0.000
12	7	0.72	3.68	2.96	0.000
13	4	0.96	3.28	2.32	0.000
14	2	0.04	1.68	1.64	0.000
15	2	0.36	1.64	1.28	0.000
16	4	1.00	3.00	2.00	0.000
17	4	2.00	3.00	1.00	0.05
18	5	2.00	3.00	1.00	0.05
19	3	1.00	2.00	1.00	0.05
20	4	1.00	4.00	3.00	0.000
21	5	2.00	4.00	2.00	0.000
22	6	3.00	5.00	2.00	0.000
23	4	2.00	3.00	1.00	0.05
24	5	2.00	5.00	3.00	0.000
25	5	3.00	4.00	1.00	0.05

* p-value 0.05

Table 3 shows the mean value of detection of flaws in pre-test score is 1.349 compared to post-test score i.e. 3.442.

Table 3: Comparison of Pre- and post-test scores for item flaws and errors

Comparison	n	Mean ± SD	No. of Questions	Sig Value
Pre-Test Score	50	1.3488 ± .78685	25	0.003
Post-Test Score		3.4416 ± .83845	25	

Table 4 represents the exercise to detect the correct cognitive level of given MCQs by the participants. The pre- and post-test questionnaire comprises 12 MCQs in which faculty were assigned to correctly identify the cognitive levels (C1, C2, C3) according to Bloom taxonomy. By comparing their results, it is revealed that in pre-test the percentage of participants who perceive correct cognitive level is 77.5% compared to 87.5% participants who detect correct cognitive level in post-test. This table also shows individual question wise analysis of detecting correct cognitive level in pre-test & post-test. It is evident from the results that participants identified the cognitive level more accurately, especially at C3 level in comparison to C1 or C2 level, after the training in post-test.

Table 4: Detection of correct cognitive level on MCQs (n=50)

Question #	Item with cognitive levels	Pre-test score	Post-test score	p-value
		% Identified	% Identified	
1	C1	70%(35)	74%(37)	0.072
2	C1	74%(37)	84%(42)	
3	C1	80%(40)	94%(47)	
4	C1	90%(45)	92%(46)	
Mean - Cognitive level 1 (C1)		78%(39)	86%(43)	
5	C2	78%(39)	80%(40)	0.135
6	C2	76%(38)	82%(41)	
7	C2	80%(40)	86%(43)	
8	C2	68%(34)	90%(45)	
Mean - Cognitive level 2 (C2)		76%(38)	84%(42)	
9	C3	90%(45)	96%(48)	0.026
10	C3	68%(34)	90%(45)	
11	C3	80%(40)	92%(46)	
12	C3	72%(36)	86%(43)	
Mean - Cognitive level 3 (C3)		78%(39)	92%(46)	
Mean of all cognitive levels		78%(39)	88%(44)	

The Table 5 describes responses obtained from participants after reviewing a post hoc analysis report consisting of details of each item showing values of difficulty index, discrimination index, and distractors analysis provided to them as a post hoc report. The participants in these 12 items were given a task to

DOI: https://doi.org/10.54393/pjhs.v4i04.700

accurately identify those items which need revision or rejected or retained based on their observation of the item analysis report. For all participants the accuracy to identify the items which need to be rejected from the questionnaire came out to be 71.4% and accuracy of participants to mark items for further improvement was 60%. With respect to the quantity of item analysis instruction they received, the results in this area were moderate.

Table 5: Accuracy of correctly identifying items from item analysis report

Items which need to be discarded according to the key	Total # of items identified by the participants which need to be discarded	Accuracy of detection in percentage	Total # items which need improvement according to the key	Total # of items identified by the participants which need improvement	Accuracy of detection in percentage
7	5	71.43%	5	3	60.00%

DISCUSSION

Prior research demonstrated that medical faculties with formal training can be more effective in their professions. Typically, teachers create the test questions themselves, though occasionally they use item test banks as a source to create new items. Faculty can develop their skills in creating effective MCQs by participating in writing workshops for MCQ items in a positive environment [17]. Our findings demonstrate the efficacy of writing workshops for MCQs was beneficial in improving the mean scores and outcome related to questions. The improvement between pre- and post-training scores shows a considerable positive impact on the participants' capacity to produce high-quality test items. After training, items flaws identified showed statistically substantial increases in mean item quality scores. Study by Nemec and Welch also signifies the improvement in quality of MCQS through faculty development seminars [18]. MCQ quality was observed to have improved by Al-Faris *et al.*, after a one-day session [19]. The value of the workshop is exemplified by the fact based on the difference in score of almost 50% in the quality of the questions on the pre-test and post-test in our study. This is emphasized by Dellinges and Curtis in their study, which found that a brief, one-hour training session for dental school faculty members resulted in considerable increases in the quality of internal MCQ item-writing [20]. Major shortcomings included a lack of homogeneity, an incorrect lead-in, improbable distractors, a clinical vignette with insufficient data, options that weren't alphabetized, and the use of negative stem. In a study by Rush *et al.*, odd stem construction, improbable distractors, and longest response is correct were the most prevalent errors in item writing [21]. Local study revealed flaws most frequently observed were, conflicting and confusing information in the stem; a lack of a clinical scenario; and an evaluation of simple recall items [22]. All of the above, none of the above and unfocused questions were less frequent in our study which is commonly reported by medical educationist as a common item writing flaws [23]. The annual flaw rates ranged from 21% in 2011 to 67% in 2009. Item flaws not only cause systemic mistakes but can harm medical students. Because of these flaws, some students may find it simpler to respond to a question accurately based on their test-taking abilities rather than their knowledge base. On the other hand some students might have difficulty in answering such MCQS just because of lack of clarity which confuses the students' understanding. Pre- and post-workshop assessments of the participants' cognitive capabilities revealed significant improvements in all three levels of cognition, but the greatest improvements were shown in higher complexity questions (C3). It might be because recall-level MCQs are simpler to make and require less time and information than problem-solving MCQs, which demand for experience and training. Abdulghani *et al.*, revealed significant improvement in mean score for high cognitive questions after a longitudinal faculty training [24]. Tenzin *et al.*, produced similar results [25]. Following the session, there was a noticeable improvement in knowledge and the participant's capacity to identify and correct errors. Besides pointing out item flaws, the majority of the questions (70%) that needed to be discarded and 60% that needed improvement were also selected by the participants when they were given sets of questions (pre- and post-workshop) with specified standards of difficulty and discriminating index. Such workshops demonstrate a considerable improvement after focused training in item writing as our faculty is usually busy in academic responsibilities and are more involved in teaching rather than assessing the students. These sessions will assist the faculty in identifying components that are improperly designed and focusing their attention on optimizing them to raise the standard of the question bank [26]. The methodology used for this research divided the 'item writing' into three different areas of research. This division provided detailed comparison between untrained and trained faculty. The sample size was limited to only 103 participants. Only immediate effect of workshop was taken into consideration. Application of this focused training on 'item writing' must be incorporated in their real life practices. Regular analysis of 'items' should be done to determine the long term effect of training programs for faculty. This will help to find out areas for improvement in assessment especially item writing.

DOI: https://doi.org/10.54393/pjhs.v4i04.700

CONCLUSIONS

The faculty training for MCQ development had a significant effect on the producing high quality MCQs. Our study revealed that workshop helped faculty identify item flaws with significant improvement in pre- and post-test scores. The cognitive capabilities of faculty were improved relatively in all three levels of cognition, but the greatest improvements were shown in higher complexity questions (C3). Besides identifying item flaws there was noticeable improvement in knowledge and the participant's capacity to identify and correct errors. Item analysis is a useful tool for identifying subpar MCQs. It enabled us to identify badly structured MCQs and focused on refining them to raise the standard of the question bank. Hence the learning and performance of medical educators in the area of developing MCQs appear to be significantly improved by such faculty development workshops.

Authors Contribution

Conceptualization: MFR
Methodology: SK
Formal analysis: SQB, SK
Writing-review and editing: MFR, SQB, TA, MA

All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Source of Funding

The authors received no financial support for the research, authorship and/or publication of this article.

REFERENCES

[1] Lockyer J, Carraccio C, Chan MK, Hart D, Smee S, Touchie C, *et al*. Core principles of assessment in competency-based medical education. Medical Teacher. 2017 Jun; 39(6): 609-16. doi: 10.1080/0142159X.2017.1315082.

[2] Schuwirth LW and van der Vleuten CP. A history of assessment in medical education. Advances in Health Sciences Education. 2020 Dec; 25(5): 1045-56. doi: 10.1007/s10459-020-10003-0.

[3] Farooqui F, Saeed N, Aaraj S, Sami MA, Amir M. A Comparison Between Written Assessment Methods: Multiple-choice and Short Answer Questions in End-of-clerkship Examinations for Final Year Medical Students. Cureus. 2018 Dec; 10(12): e3773. doi: 10.7759/cureus.3773.

[4] Holzinger A, Lettner S, Steiner-Hofbauer V, Capan Melser M. How to assess? Perceptions and preferences of undergraduate medical students concerning traditional assessment methods. BMC Medical Education. 2020 Sep; 20(1): 312. doi: 10.1186/s12909-020-02239-6.

[5] Adnan S, Sarfaraz S, Nisar MK, Jouhar R. Faculty perceptions on one-best MCQ development. The Clinical Teacher. 2023 Feb; 20(1): e13529. doi: 10.1111/tct.13529.

[6] Balaha MH, El-Ibiary MT, El-Dorf AA, El-Shewaikh SL, Balaha HM. Construction and Writing Flaws of the Multiple-Choice Questions in the Published Test Banks of Obstetrics and Gynecology: Adoption, Caution, or Mitigation? Avicenna Journal of Medical Sciences. 2022 Aug; 12(3): 138-47. doi: 10.1055/s-0042-1755332.

[7] Arooj M, Mukhtar K, Khan RA, Azhar T. Assessing the educational impact of cognitive level of MCQ and SEQ on learning approaches of dental students. Pakistan Journal of Medical Sciences. 2021 Mar; 37(2): 445-9. doi: 10.12669/pjms.37.2.3475.

[8] Pickford JC and Newcomb LH. Relationship of cognitive level of instruction to students' cognitive level of achievement. NACTA Journal. 1989 Jun; 33(2): 56-9.

[9] Zaidi NB, Hwang C, Scott S, Stallard S, Purkiss J, Hortsch M. Climbing Bloom's taxonomy pyramid: Lessons from a graduate histology course. Anatomical Sciences Education. 2017 Sep; 10(5): 456-464. doi: 10.1002/ase.1685.

[10] Stringer JK, Santen SA, Lee E, Rawls M, Bailey J, Richards A, *et al.* Examining Bloom's Taxonomy in Multiple Choice Questions: Students' Approach to Questions. Medical Science Education. 2021 May; 31(4): 1311-7. doi: 10.1007/s40670-021-01305-y.

[11] Khan MU and Aljarallah BM. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. International Journal of Health Sciences. 2011 Jan; 5(1): 39-43.

[12] Grainger R, Dai W, Osborne E, Kenwright D. Medical students create multiple-choice questions for learning in pathology education: a pilot study. BMC Medical Education. 2018 Aug; 18(1): 201. doi: 10.1186/s12909-018-1312-1.

[13] Kowash M, Hussein I, Al Halabi M. Evaluating the Quality of Multiple Choice Question in Paediatric Dentistry Postgraduate Examinations. Sultan Qaboos University Medical Journal. 2019 May; 19(2): e135-e141. doi: 10.18295/squmj.2019.19.02.009.

[14] Islam ZU and Usmani A. Psychometric analysis of Anatomy MCQs in Modular examination. Pakistan Journal of Medical Sciences. 2017 Sep; 33(5): 1138-43. doi: 10.12669/pjms.335.12382.

[15] Sajjad M, Iltaf S, Khan RA. Nonfunctional distractor analysis: An indicator for quality of Multiple choice

DOI: https://doi.org/10.54393/pjhs.v4i04.700

questions. Pakistan Journal of Medical Sciences. 2020 Jul; 36(5): 982-6. doi: 10.12669/pjms.36.5.2439.

[16] Ali SH, Carr PA, Ruit KG. Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter. Journal of the Scholarship of Teaching and Learning. 2016 Feb; 16(1): 1-4. doi: 10.14434/josotl.v16i1.19106.

[17] Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK, Syed S, *et al*. Faculty development programs improve the quality of Multiple Choice Questions items' writing. Scientific Reports. 2015 Apr; 5(1): 1-7. doi: 10.1038/srep09556.

[18] Nemec EC and Welch B. The impact of a faculty development seminar on the quality of multiple-choice questions. Currents in Pharmacy Teaching and Learning. 2016 Mar; 8(2): 160-3. doi: 10.1016/j.cptl.2015.12.008.

[19] AlFaris E, Naeem N, Irfan F, Qureshi R, Saad H, Al Sadhan RE, *et al*. A one-day dental faculty workshop in writing multiple-choice questions: an impact evaluation. Journal of Dental Education. 2015 Nov; 79(11): 1305-13. doi: 10.1002/j.0022-0337.2015.79.11.tb06026.x.

[20] Dellinges MA and Curtis DA. Will a short training session improve multiple-choice item-writing quality by dental school faculty? A pilot study. Journal of Dental Education. 2017 Aug; 81(8): 948-55. doi: 10.21815/JDE.017.047.

[21] Rush BR, Rankin DC, White BJ. The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. BMC Medical Education. 2016 Dec; 16(1): 1-10. doi: 10.1186/s12909-016-0773-3.

[22] Rauf AR and Sultana S. Effect of faculty training on quality of Multiple Choice Questions. Rawal Medical Journal. 2021 Dec; 46(2): 430.

[23] Hingorjo MR and Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. Journal of Pakistan Medical Association. 2012 Feb; 62(2): 142-7.

[24] Abdulghani HM, Irshad M, Haque S, Ahmad T, Sattar K, Khalil MS. Effectiveness of longitudinal faculty development programs on MCQs items writing skills: A follow-up study. PloS One. 2017 Oct; 12(10): e0185895. doi: 10.1371/journal.pone.0185895.

[25] Tenzin K, Dorji T, Tenzin T. Construction of multiple choice questions before and after an educational intervention. Journal of Nepal Medical Association. 2017 Jan; 56(205): 112-6. doi: 10.31729/jnma.2976.

[26] Gupta P, Meena P, Khan AM, Malhotra RK, Singh T. Effect of faculty training on quality of multiple-choice questions. International Journal of Applied and Basic Medical Research. 2020 Jul; 10(3): 210. doi: 10.4103/ijabmr.IJABMR_30_20.