Original Article

Effect of Faculty Training on Quality of Multiple-Choice Questions

Piyush Gupta, Pinky Meena[1], Amir Maroof Khan[2], Rajeev Kumar Malhotra[3], Tejinder Singh[4]

Departments of Pediatrics and [2]Community Medicine, Medical Education Unit, University College of Medical Sciences, [1]Department of Pediatrics, University College of Medical Sciences, [3]Delhi Cancer Registry, Dr. BRA Institute Rotary Cancer Hospital, AIIMS, Delhi, [4]Department of Pediatrics and Medical Education, SGRD Institute of Medical Sciences and Research, Amritsar, Punjab, India

Abstract

Background: Multiple-choice question (MCQ) is frequently used assessment tool in medical education, both for certification and competitive examinations. Ill-constructed MCQs impact the utility of the assessment and thus the fate of examinee. We conducted this study to ascertain whether a short training session for faculty on MCQ writing results in desired improvement in their item-writing skills. **Methods:** A 1-day workshop on constructing high-quality MCQs was conducted for the faculty as a before-after design, following training session of 3 h duration. 28 participants wrote preworkshop (n = 133) and postworkshop (n = 137) MCQs, which were analyzed and compared for 17 item-writing flaws. A mock test of 100 MCQs (selected by stratified random sampling from all the MCQs generated during the workshop) was conducted for MBBS-passed students for item analysis. **Results:** Item-writing flaws reduced following the training (15% vs. 27.7%, $P < 0.05$). Improvement mainly occurred in quality of options; heterogeneity dropped from 27.1% prior to the workshop to 5.8% postworkshop. The proportion of MCQs failing the cover test remained similarly high (68.4% vs. 60.6%), and there was no improvement in writing of the stem before and after the workshop. The item analysis did not reveal any significant improvement in facility value, discriminating index, and proportion of nonfunctioning distractors. **Conclusion:** A single, short-duration faculty training session is not good enough to correct flaws in writing of the MCQs. There is a need of focused training of the faculty in MCQ writing. Courses with a longer duration, supplemented by repeated or continuous faculty development programs, need to be explored.

Keywords: *Faculty training, multiple-choice questions, quality of multiple-choice questions*

Introduction

Multiple-choice question (MCQ) format is one of the most common tools of assessment used at almost all levels and across all specialties, both for formal certification and competitive examinations. They can assess factual recall, problem-solving, and reasoning. Good-quality MCQs instigate critical thinking requiring interpretation, integration, synthesis, and analysis of the medical knowledge and facts. This mode of assessment therefore determines students' learning behavior. They are preferred for their objectivity and ease of scoring a large bulk of students at a time.[1-3] Along with ensuring fairness, well-constructed MCQs can differentiate between high-performing and low-performing students. Correct framing of MCQ (also known as item) is thus essential.

The crucial task of item formulation is ascertained to the multitasking faculty members. Most faculty members are either not acquainted to standard MCQ guidelines or refrain from changing practice. Vyas and Supe reported that flaws in writing MCQ are primarily because of lack of faculty training.[4] Abdulghani *et al*. from Riyadh reported significant improvement in framing of MCQs following a faculty development program (FDP).[5] Similar results have been shown by Naeem *et al*.[6] and Tenzin *et al*. (2017).[7] However, there is lack of research on this topic in Indian context.

The Medical Council of India (MCI) has made it mandatory for the faculty in medical colleges in India to attend the revised Basic Course Workshop (rBCW) in medical education. The workshop also has a training session on MCQs. We conducted this study to ascertain whether the training session on MCQ, as conducted in rBCW, results in desired improvement in item-writing skills of the faculty. The aim of the study was to evaluate the efficacy of a faculty development workshop. The objectives were to compare the item-writing flaws (IWFs), Facility

How to cite this article: Gupta P, Meena P, Khan AM, Malhotra RK, Singh T. Effect of faculty training on quality of multiple-choice questions. Int J App Basic Med Res 2020;10:210-4.

Submitted: 18-Jan-2020
Revised: 25-Feb-2020
Accepted: 26-Apr-2020
Published Online: 11-Jul-2020

Address for correspondence:
Dr. Piyush Gupta, Department of Pediatrics, University College of Medical Sciences, Delhi - 110 095, India. E-mail: prof.piyush.gupta@gmail.com

Access this article online

Website: www.ijabmr.org

DOI: 10.4103/ijabmr.IJABMR_30_20

Quick Response Code:

Value (FV), discriminating index (DI), and nonfunctioning distractors (NFDs), before and after the workshop.

Methods

The study was conducted from October 2018 to March 2019. It was an interventional study with pre–post design to test the efficacy of intervention, i.e., faculty training session for framing MCQs. The study protocol was discussed in Medical Education Unit of the institute, for fine-tuning the methodology. Ethical clearance was obtained from the Institutional Ethical Committee (IEC), and written informed consent was obtained from the participating faculty and students.

Participation was invited from among the faculty members of the University College of Medical Sciences to attend the 1-day MCQ workshop. The circular was sent by the coordinator of the Medical Education Unit (MEU) in mid-November and those who have intention to attend were invited by December 20, 2018. A poster was also prepared by the MEU and widely circulated to all the notice boards of various departments and other prominent places in the college and hospital. This was also circulated by E-mail on social media groups of the faculty and shared on website of the MEU. After working on the logistics and developing adequate tools of the workshop, it was conducted on February 04, 2019. An external expert from the Regional Center for Faculty Development at the Maulana Azad Medical College was invited to conduct the workshop. The enrolment was planned on first-come, first-served basis. We planned to enroll between 25 and 30 participants for the workshop. Those finally enrolled were asked to provide consent as per IEC requirements and attend the 1-day session on the designated date. A meeting of the MEU was held a month prior to the workshop and all logistics were discussed and work distributed for the workshop.

Intervention

The workshop was held on February 04, 2019. The agenda was circulated to the participant faculty well in advance by E-mail. A subject-wise list of participants was prepared according to the applications received. We procured almost all the standard textbooks used by the undergraduate students for these subjects from library and other sources. These were made available to the participants during the workshop. All the participants were given a book each from their respective subjects and asked to prepare 4–5 MCQs (MCQ stem with 4 choices and single best response), targeted at NEET PG entrance examination for MBBS passed students. One hour was allotted for this task. MCQs such generated were collected and labeled as preworkshop MCQs. This was followed by the teaching–learning (T-L) session and workshop for next 3 h. The T-L session was conducted as per the MCI regional center guidelines laid down for framing of MCQs in the rBCW for medical faculty. Following the workshop, the participants were again asked to frame another 4–5 MCQs in their respective subject, based on the skills gained and directives received during the workshop. One hour was allotted for this activity. Postworkshop MCQs were also collected. Finally, a feedback was obtained from the participants regarding their satisfaction about the workshop.

For data analysis, all the pre- and postworkshop MCQs thus generated were complied, typed, proofread, and analyzed for the IWFs (16 criteria). MCQs with IWFs are those items which fulfill one or more of the criteria given in Table 1.

Further, an item analysis was conducted on 100 questions selected by stratified random sampling from all the MCQs generated during the workshop. The first stratification was done to select 50 MCQs each from the pre- and postworkshop pools. The second stratification was based on having equal representations of all the participants and subjects in each of two groups (i.e., pre- and postworkshop questions). These 100 questions were then mixed up to prepare a final question paper, ready for administration. The key to all the questions was prepared and stored separately. The question paper was administered to 21 interns who were preparing for NEET PG entrance test and voluntarily agreed to undergo the mock exam. The answers were checked, marked, and then arranged in rank order, with student scoring highest marks at the top, for each question separately. The MCQs were assessed on the following outcome indicators: difficulty index, discrimination index, and nonfunctioning distracter. These outcome measures are defined and discussed as follows.

1. Facility Value (FV): It indicates the percentage of students who correctly answered a given test item. Easier the item, higher is its FV. From a score of 0–100, FV of 70% indicates easy; 20%–70% indicates moderate; and <20% indicates difficult test item.[5] Moderate difficulty items (20%–70%) in a test have better discriminating ability. It is effected by the cognitive level of the question, the content of the stem, and adequate number of plausible options[8]
2. Discrimination index (DI): It is the ability of a test item to discriminate between high (top 30%) and low examinee (bottom 30%) scorers. Higher the discriminating indices of a test item, better is its discriminating capability. The cutoff values for the discrimination index (DI) were taken as DI >0.15 and non-DI ≤0.15.[9] This was calculated by the following formula:
   DI = 2 × (HAG − LAG)/N (HAG refers to high scorers, and LAG low scorers).
3. NFD (%): The options of a test item that have been selected by less than 5% of the examinees are called NFD.[10] These options are often unrelated or quite easy to be figured out by simple guesswork. The ineffective options change the difficulty level of the question and affect the discriminating ability of the test item.

Table 1: Comparison of item-writing flaws in pre-and postworkshop multiple-choice questions

	Item-writing flaw	Preworkshop (*n*=133), *n* (%)	Postworkshop (*n*=137), *n* (%)
1	Negative questions. e.g., Which ONE of the following is NOT a characteristic…	9 (6.8)	10 (7.3)
2	"EXCEPT" marked question	32 (24.1)	9 (6.6)*
3	Usage of ambiguous (e.g., frequently, often, occasionally) or absolute terms (e.g. almost, never, frequent)	11 (8.3)	4 (2.9)*
4	A word in the stem repeated in the option(s)	-	1 (0.70)
5	Options are not uniform/heterogeneous	36 (27.1)	8 (5.8)*
6	Has a single long option, which is the correct answer	2 (1.5)	3 (2.2)
7	Uses "all of the above" or "none of the above" option	4 (3.0)	3 (2.2)
8	Options are not in a chronological order	6 (4.5)	4 (2.9)
9	There is more than one correct answer	3(2.3)	-
10	Options are overlapping	4 (3.0)	-
11	Stem is not clear	25 (18.8)	33 (24.1)
12	Answer is hinged to another question	-	-
13	Is a true or false type of question	7 (5.3)	4 (2.9)
14	Case-scenario, if used; is not related to the question	-	-
15	Does not fulfil the cover test, i.e., if you cover the options, you cannot answer the questions	91 (68.4)	83 (60.6)
16	Key not provided	18 (13.5)	29 (21.2)

*$P<0.05$

Statistical analysis

Data were entered into Excel and analyzed by SPSS Version 25 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean (standard deviation) and categorical variables were expressed in numbers and proportions. Chi-square/Fisher's exact test were used to compare the categorical variables between the pre- and postworkshop MCQs. For comparing the continuous variables, nonparametric tests of significance were used as the data did not follow a normal distribution. $P < 0.05$ was considered as significant.

Results

Overall, 112 faculty were approached, of whom 25 showed their willingness to attend the workshop. 8 faculty members were also enrolled from other medical colleges of Delhi. Of 33 participants thus enrolled, 28 finally participated in the workshop. The participants ranged from assistant professors to professors with 3–30 years (median 10 years) teaching experience. The subject representation of the participating faculty delegates was as follows: anesthesia (3), biochemistry (3), community medicine (4), internal medicine (2), microbiology (1), obstetrics (2), otolaryngology (1), pathology (4), pediatrics (3), pharmacology (2), physiology (2), and surgery (1).

A total of 133 questions were generated in the preworkshop session and 137 questions were generated in the postworkshop session. There were 20/133 (15%) MCQs without any flaw in the preworkshop questions, as compared to 38/137 (27.7%) questions without any flaw in the postworkshop questions ($P = 0.01$). Table 1 compares the proportions of individual IWFs before and after the workshop. There were 251 flaws in the preworkshop questions compared to 191 in the postworkshop questions. Statistically significant improvement occurred only in two of the 16 flaws. Frequency of 3 flaws increased after the workshop; however, it was not statistically significant. Table 2 presents a cross tabulation of the number of faulty MCQs with frequency of IWFs, before and after the workshop.

Pre- and postworkshop questions were found to be equally difficult or easy as assessed by the item analysis. The mean facility value in the two groups was also not statistically different ($P = 0.81$). NFDs were present in an almost equal proportion of items written before (31/150, 21%) or (39/150, 26%) after the workshop. A total of 35 (70%) of items were able to discriminate between high and low scorers before the workshop as compared to 32 (64%) after the workshop. The difference was not found to be statistically different between pre- and postworkshop test items ($P = 0.52$) [Table 3].

Discussion

Overall, the quality of MCQs prepared by experienced faculty was unsatisfactory. We observed a statistically significant reduction in frequency of IWFs following the training of faculty in writing MCQs. Prior to training, only 15% of the MCQs prepared were flawless. This increased to 27.7% after the workshop. Though the improvement was statistically significant, it was not sufficient enough to have an educational impact. Improvement mainly occurred in the selection of options which were more homogeneous after the workshop. As compared to 27.1% heterogeneity in options prior to the workshop, the frequency dropped to

Table 2: Comparison of number of multiple-choice questions with frequency of item-writing flaws in pre- and postworkshop multiple-choice questions

Number of item-writing flaws	Number of MCQs with flaws	
	Preworkshop (*n*=113)	Postworkshop (*n*=99)
1	30	36
2	44	42
3	27	17
4	6	4
5	6	Nil

MCQ: Multiple-choice question

Table 3: Comparison of item analysis before and after workshop

Item analysis parameter	Preworkshop (*n*=50)	Postworkshop (*n*=50)	*P*
Facility value, *n* (%)			
<20	12 (24)	12 (24)	0.89
20-70	28 (56)	30 (60)	
>70	10 (20)	8 (16)	
Mean (SD)	42.1 (26.6)	43.3 (24.9)	0.81
Discriminating index >0.15, *n* (%)	35 (70)	32 (64)	0.52
Nonfunctioning distractors (*n*=150), *n* (%)	31 (21)	39 (26)	0.85

SD: Standard deviation

5.8%. No improvement was noticed in writing the stem. Furthermore, the proportion of MCQs failing the cover test remained similarly high, before and after the session (68.4% vs. 60.6%). Item analysis also did not reveal any significant improvement in FV, DI and proportion of NFDs ($P > 0.05$) after the workshop.

FDPs in MCQ writing holds promise and earlier studies have shown significant improvement following focused training in item writing.[5-7] Better test outcomes are reported after longitudinal and repeated training sessions. Abdulghani *et al*. reported significant improvement in FV and DI increased distractor efficiency (DE) mean score and high cognitive level questions during each successive academic year, after longitudinal faculty training. Easy and poor discriminating questions, NFDs and IWFs were decreased significantly. They conducted 1-day long training workshop twice in an academic year. Overall impact on student competency and learning was positive.[11] There is a school of thought who propagates development of dedicated FDPs where focus is on high quality content, practice, feedback, and improvement. Involvement of expert medical educationist is pitched for the same.[12,13]

Longitudinal FDPs are resource intensive apart from committing longer duration of faculty time. Short-duration FDPs like we conducted and also that conducted in rBCW have not been adequately evaluated, in terms of functional output. These sessions evaluate pre- and post-training scores and lack long-term follow-up.[14] However, repeated short-course training has shown improved test preparation. Initial 3 h session followed by two, 2 h sessions over a period of 3 months was conducted as a study. This breakdown and repetition of high-quality sessions containing item analysis and discussion of the feedback resulted in improved item-writing skills in faculty.[15] Al-Faris observed improved quality of MCQ, following a 1-day training.[16]

Our study did not show much improvement in item writing. This could be attributed to shorter duration of the session, with inadequate time for hands-on exercise and critical analysis and reviewing of test items prepared by the participants. Our study showed almost equal number of unclear stem, similar discrimination index, and NFDs pre- and postworkshop MCQs. Constructing plausible distractors is difficult and demands clarity in objective of the question. Tarrant *et al*. opined that in most cases, the number of plausible distractors should be three.[10] Clear objectives with deep learning stimulating content of the test items ensure reliability of test and improve competence and student learning behavior.[17] Often untrained faculty write items with options that are just fillers and have negative impact on discriminating ability of the MCQ.

The study had several limitations. The selection bias for the participants could not be ruled out. The group for training was heterogeneous in terms of their experience in medical education. The teaching–learning process was centered more on deficiencies rather than highlighting what is correct. However, the content of teaching–learning session and resource person were same as delivered in rBCW of MCI for the faculty. We feel that the teaching–learning session needs restructuring. Training can be conducted by several faculty members over a longer time. Slots should be created for more hands-on exercises and one-to-one interaction. Item flaws should be identified immediately, and all participants should be given the chance to improve them in the workshop itself. Future studies can focus on comparison of the two methods of conducting this workshop by assessing their impact.

Conclusion

Though from a limited sample, it appears that most Faculty in Medical Colleges is ill-equipped to write flawless MCQs for MBBS level examinations. A traditional faculty training of 3 h (as imparted in the MCI Basic Course Workshop) is inadequate for causing the desired improvement in the quality of MCQs as assessed by nonimprovement in IWFs or item-analysis. Courses with a longer duration, supplemented by repeated or continuous FDPs, need to be explored.

What this study adds

Single short duration faculty training session is not good enough to correct flaws in writing of the MCQs.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Tarrant M, Ware J. A framework for improving the quality of multiple-choice assessments. Nurse Educ 2012;37:98-104.
2. Abdel-Hameed AA, Al-Faris EA, Alorainy IA, Al-Rukban MO. The criteria and analysis of good multiple choice questions in a health professional setting. Saudi Med J 2005;26:1505-10.
3. Case SM, Swanson DB. Constructing Written Test Questions for the Basic and Clinical Sciences. 3rd ed. Philadelphia: National Board of Medical Examiners; 2002.
4. Vyas R, Supe A. Multiple choice questions: A literature review on the optimal number of options. Natl Med J India 2008;21:130-3.
5. 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. Sci Rep 2015;5:9556.
6. Naeem N, van der Vleuten C, Alfaris EA. Faculty development on item writing substantially improves item quality. Adv Health Sci Educ Theory Pract 2012;17:369-76.
7. Tenzin K, Dorji T, Tenzin T. Construction of multiple choice questions before and after an educational intervention. JNMA J Nepal Med Assoc 2017;56:112-6.
8. Abdulghani HM, Ponnamperuma G, Ahmad F, Amin Z. A comprehensive, multi-modal evaluation of the assessment system of an undergraduate research methodology course: Translating theory into practice. Pak J Med Sci 2014;30:227-32.
9. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc 2012;62:142-7.
10. 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.
11. 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;12:e0185895.
12. Ebrahimi S, Kojuri J. Assessing the impact of faculty development fellowship in Shiraz University of Medical Sciences. Arch Iran Med 2012;15:79-81.
13. Singh T, de Grave W, Ganjiwale J, Supe A, Burdick WP, van der Vleuten C. Impact of a fellowship program for faculty development on the self-efficacy beliefs of health professions teachers: A longitudinal study. Med Teach 2013;35:359-64.
14. Dellinges MA, Curtis DA. Will a short training session improve multiple-choice item-writing quality by dental school faculty? A pilot study. J Dent Educ 2017;81:948-55.
15. Iramaneerat C. The impact of item writer training on item statistics of multiple-choice items for medical student examination. Siriraj Med J 2012;64:178-82.
16. AlFaris E, Naeem N, Irfan F, Qureshi R, Saad H, Al Sadhan R, *et al*. A one-day dental faculty workshop in writing multiple-choice questions: An impact evaluation. J Dent Educ 2015;79:1305-13.
17. Collins J. Education techniques for lifelong learning: Writing multiple-choice questions for continuing medical education activities and self-assessment modules. Radiographics 2006;26:543-51.