Impact of a Longitudinal Faculty Development Program on the Quality of Multiple-Choice Question Item Writing in Medical Education

Lukman Femi Owolabi¹, Bappa Adamu¹, Magaji Garba Taura², Adamu Imam Isa², Abubakar Muhammed Jibo⁴, Reda Abdul-Razek¹, Mufarrah Muhammed Alharthi⁸, Mushabab Alghamdi¹

Departments of ¹Medicine, ²Anatomy, ³Physiology, ⁴Community of Medicine and ⁵Family Medicine, University of Bisha Medical College, Bisha, Saudi Arabia

Abstract

Background: Like many other academic programs, medical education is incomplete without a robust assessment plan. Objective: The study aimed to evaluate the impact of longitudinal faculty development program (FDP) on the examination item quality (EIQ) from a cohort of medical college faculty members. Methods: Item analysis (IA) of multiple-choice questions (MCQs) from a cohort of medical tutors over a 3-year period (2017 [S1], 2018 [S2], and 2019 [S3]) before and following once-per-week FDP was conducted. The questions were from three randomly selected courses: man and his environment (MEV) from phase 1, central nervous system (CNS) from phase 2, and internal medicine (MED) from phase 3. Data assessed were 480 MCQs from the final exams in the courses. The parameters considered in IA were the difficulty index, index of discrimination, nonfunctional distractors (NFDs), distractor efficiency for each question item, and Cronbach’s alpha (CA) for the test as a whole. Comparison over the 3 years was made using Fisher’s exact test and repeated-measures ANOVA with Bonferroni test as post hoc test. Results: Overall, out of 480 MCQs, 272 had no NFD (52 [19.52%], 104 [38.24%], and 116 [42.65%] in 2017, 2018, and 2019, respectively) with a significant difference between S3, S2, and S1 (P < 0.0001). The mean CA for the exams in S1, S2, and S3, respectively, were 0.51, 0.77, and 0.84, P < 0.0001. Conclusion: There was an improvement in EIQ following the implementation of longitudinal FDP. Thus, the need for active training and retraining of the faculty for a better EIQ cannot be overemphasized.

Keywords: Faculty development, impact, multiple-choice question, medical education

Résumé

Contexte: Comme beaucoup d’autres programmes universitaires, la formation médicale est incomplète sans un plan d’évaluation solide. Objectif: L’étude visait à évaluer l’impact du programme longitudinal de formation professorale (FDP) sur la qualité des éléments d’examen (EIQ) d’une cohorte de membres du corps professoral des facultés de médecine. Méthodes: Analyse des éléments (IA) des questions à choix multiples (QCM) d’une cohorte de tuteurs médicaux une période de trois ans (2017 [S1], 2018 [S2] et 2019 [S3]) avant et après le déroulement du FDP hebdomadaire a été effectuée. Les questions venaient de trois cours choisis au hasard: l’homme et son environnement (MEV) de la phase 1, le système nerveux central (SNC) de la phase 2 et interne médecine (MED) de la phase 3. Les données évaluées étaient 480 QCM des examens finaux des cours. Les paramètres considérés dans IA étaient l’indice de difficulté, l’indice de discrimination, les distracteurs non fonctionnels (NFD), l’efficacité du distracteur pour chaque question et le alpha (CA) pour le test dans son ensemble. La comparaison au cours des 3 années a été faite en utilisant le test exact de Fisher et l’ANOVA à mesures répétées avec Test de Bonferroni comme test post hoc. Résultats: Dans l’ensemble, sur 480 QCM, 272 n’avaient pas de NFD (52 [19,52%], 104 [38,24%] et 116 [42,65%] en 2017, 2018 et 2019, respectivement) avec une différence significative entre S3, S2 et S1 (P<0,0001). L’AC moyenne pour les examens en S1, S2

Access this article online

Quick Response Code:

Website: www.annalsafrmed.org

DOI: 10.4103/aam.aam_14_20

Address for correspondence: Dr. Lukman Femi Owolabi, Department of Medicine, University of Bisha Medical College, Bisha, Saudi Arabia. E-mail: drlukmanowolabi@yahoo.com

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKLHRPMedknow_reprints@wolterskluwer.com

How to cite this article: Owolabi LF, Adamu B, Taura MG, Isa AI, Jibo AM, Abdul-Razek R, et al. Impact of a longitudinal faculty development program on the quality of multiple-choice question item writing in medical education. Ann Afr Med 2021;20:46-51.
et S3, respectivement, était de 0,51, 0,77 et 0,84, P <0,0001. Conclusion: Il y a eu une amélioration de l’EIQ après la mise en œuvre du FDP. Ainsi, la nécessité d’une formation active et d’un recyclage de la faculté pour un meilleur QEI ne peut pas être surestimée.

Mots-clés: Développement du corps professoral, impact, question à choix multiples, formation médicale

**Introduction**

The context in which students learn has a strong influence on how they learn.\(^1\) The way the learning is organized, facilities that are available for learning, and assessment method chosen for the evaluation of their learning have been identified as the pivotal factors central to an effective learning.\(^\[1\]\)

All over the world, students’ learning is, to a large extent, enhanced and driven by assessment.\(^{1,2}\) In medical education, like in other forms of education, assessment plays a very crucial role in giving the teachers feedback on their students’ educational activities, thus, the quality of test questions aimed at assessing students’ learning achievement remains a very critical issue in medical education. Consequently, development of standardized and qualitative test skills becomes not only necessary but also imperative for tutors in medical schools.\(^3\) More often than not, tutors develop examination questions by themselves or sometimes rely on questions stored in question banks for the students’ examination. In the course of developing or recalling questions, unforeseen errors could occur. Such errors could, among other reasons, be due to a dearth of background knowledge on test item development.

Since the mid-20th century, multiple-choice questions (MCQs) have gained popularity in institutions around the world as a competency test for objectively assessing students’ knowledge and understanding and application of such knowledge in addition to its ability to test the higher levels of cognition.\(^4\)\(^,\)\(^5\) The vintage position of MCQ in the field of assessment is ascribed to its higher reliability, validity, fairness, better flexibility, variety, ease of creation, and ease of scoring.\(^6\) Other things that make MCQs appealing to teachers, particularly those who teach a large number of students, is the fact that they can be objectively graded and are amenable to statistical maneuver. In spite of these overwhelming characteristics, guidelines for the development of MCQ items can be easily violated when the concerned faculties are not adequately educated and professionally trained for the development of test items.\(^7\)\(^,\)\(^8\)

One of the most common errors encountered during MCQs’ quality assessment is the occurrence of item-writing flaws (IWFs) which are described as any type of violations of item-writing guidelines that can influence the performance of students on MCQs, resulting in a difficult or easy item.\(^9\) The IWFs have been reported as one of the major explanations for the dearth of quality MCQs, and that lack of or inadequacy of training in MCQ item writing significantly contributes to the flaws.\(^9\)\(^,\)\(^10\)

Generally, teachers of medical and health sciences require a background in educational theory and practice.\(^9\) A structured process that is poised to accommodate planned learning objectives and modern assessment will undoubtedly be needed for a sustainable and effective judgment on the students’ learning achievement.\(^9\)\(^,\)\(^10\)

In light of this, several academic institutions of learning have commenced in-house faculty development programs (FDPs). On the one hand, faculty development is an educational procedure aimed at preparing and enhancing the productivity of academic staff for roles that are not only germane and relevant for teaching, assessment, and research activities but are also necessary for managerial and administrative skills.\(^3\)

On the other hand, it is a veritable tool for the development and facilitation of resource material and provision of updates on instructional methods which are required for an active and effective student-centered learning.\(^11\)

Judgment as to whether a FDP is successful or not rests on the outcome of evaluation of such program. Therefore, the current study aimed to evaluate the impact of FDP on the examination item quality from a cohort of medical college faculty members in Saudi Arabia.

**Methods**

**Study setting**

University of Bisha College of Medicine (UBCOM) is a fairly new medical school in Saudi Arabia. Established in 2014 with the overriding aim of participating in the development of health care in the country, UBCOM incorporated an innovative integrated educational method in its medical education curriculum. Apart from the preparatory phase spanning a period of 1 year, the Bachelor of Medicine and Surgery degree curriculum of the college is distributed over 5 years in three phases namely the basic medical sciences, the preclinical, and the clerkship.

From its inception, UBCOM introduced some novel instructional methods such as interactive lectures (ILs), seminars, practical sessions, self-directed learning (SDL) and early clinical exposure.\(^12\) They also have clinical skill laboratory activities and a hybrid Problem-based learning (PBL).\(^12\) PBL is characterized by integration of basic sciences and clinical subjects after introducing the fundamental concepts to students in lectures, practical and skill laboratory sessions.\(^13\) The other instructional methods used in the college include team-based learning (TBL) as well as bed-side teaching and case-based learning that are introduced during the clerkship phase. After their training, the students are expected to have rotations and training in the hospital in all the required disciplines to complete the internship clinical requirements. The modes of assessment of
theoretical knowledge in UBCOM are MCQs, short-answer questions, and modified essay questions.

**Participants**

The three randomly selected courses for the study were taught and assessed by a total of 34 academic staff of UBCOM across a three-session period. The MCQs for the examinations were presented by the course coordinators to the student assessment committee (SAC), and the questions were thoroughly reviewed by the SAC based on a robust SAC policy of the college. The faculty members were statutorily required to follow the MCQ guideline in the SAC policy with emphasis on constructive alignment of assessment, that is the cognitive level of the assessment should align with that of the course objectives, specific learning objectives, instructional activities and the materials provided to students.

**Faculty development program activity and description of the selected courses**

The faculty development unit of UBCOM organizes a weekly workshop for the staff on instructional and assessment methods adopted by the college with inclusion of the Bloom’s taxonomy, MCQs, MCQ IWF, and item analysis (IA) of MCQs among others. The FDP activities were delivered in the form of workshops, seminars, and ILs. Attendance of the development program was compulsory for all the members of academic staff in the college. Knowledge acquired from these activities was further enhanced by SAC scrutiny of the questions.

The courses that were randomly selected for this study were man and his environment (MEV) from phase 1, central nervous system (CNS) from phase 2, and internal medicine from phase 3.

MEV is a 5-week and five-credit unit course. It is a broad course that encompasses many disciplines. It deals with the intercellular and intracellular environment and body’s response to internal and external insults. Faculty for the block is drawn from various participating disciplines of the college. These include anatomy, histology, physiology, biochemistry, pathology, microbiology, pharmacology, and community medicine.

Nervous system and special senses (CNS) course is an 8-week and eight-credit unit course that involves many disciplines in an integrated format. Its coverage ranges from basic medical sciences to clinical sciences. It deals with the development, structure, function, and disorders of the nervous system and special senses.

Internal medicine (MED) course is a 9-week and nine-credit unit course, in which students are exposed to many common medical conditions. Knowledge about common medical diseases affecting various organ systems including signs, symptoms, differential diagnosis, investigations, management, complications, and prevention is also discussed. In the MED, students are tutored to acquire the skills related to clinical history taking, physical examination, investigations, and treatment and procedures related to common and other important clinical conditions.

**Quality evaluation of the multiple-choice question items**

The outcome of IA in the first, second, and third sessions was used as a surrogate for the tutors’ performance in MCQ items’ writing over the three sessions under consideration. An MCQ with a single correct answer out of four alternatives was used. Apperson datalink 3000 optical reader scanner (Apperson Inc; Charlotte, NC, USA) was used to score and generate the IA figures for each of the courses.

The MCQ IA parameters considered were difficulty index (DI), index of discrimination (IOD), distractor analysis (nonfunctional distractors [NFDs] and distractor efficiency [DE]), point biserial (PB), and Cronbach’s alpha (CA). DI was defined as the percentage of students who answered a given test item correctly. DI of <0.3 was considered difficult (abnormal) and >0.7 was considered easy.

IOD is the ability of a test item to discriminate between high and low performers. IOD of <0.25 was considered abnormal. NFD is a distractor selected by <5% of the examinees. PB is a measure of item reliability. It correlates students’ scores on one particular question with their scores on their test as a whole based on the assumption that the student who scores well on a test as a whole should on average score well on the question under review. A zero or negative PB (NPB) was considered abnormal. CA is a measure of reliability and internal consistency of an examination. CA above 0.7 was considered desirable in the study.

**Statistical analysis**

The IA data obtained from the Apperson datalink 3000 optical reader scanner (Apperson Inc; Charlotte, NC, USA) were entered in a Microsoft Excel file and then transferred to STATA software, version 12.0 (Stata Corp., College Station, TX, USA). Categorical and continuous variables were described in terms of proportion and mean with standard deviation, respectively. Comparison over the three sessions (S1, S2, and S3) was made using Chi-square test, Fisher’s exact test, and repeated-measures ANOVA with Bonferroni test as post hoc test. The analysis was carried out using STATA version 12.0 (Stata Corp., College Station, TX, USA). The statistical significance level was set as $P = 0.05$ during the entire analysis.

**Ethical consideration**

Ethical approval for this study was granted by the University of Bisha Ethical Review Committee.

**Results**

Overall, 480 MCQ items with 1440 distractors were analyzed. Out of the 480 MCQs, 272 had no NFD (52 [19.52%], 104 [38.24%], and 116 [42.65%] in 2017 [S1], 2018 [S2], and 2019 [S3], respectively) with a statistically significant difference between S2 and S1 ($P < 0.0001$) [Table 1]. The post hoc analysis revealed a statistically significant improvement in DI and IOD of S3 over S2 and S1 ($P < 0.0001$ and $P = 0.0004$, respectively). The mean DE improved from...
Owolabi, et al.: Effect of faculty development program on examination item quality

Annals of African Medicine ¦ Volume 20 ¦ Issue 1 ¦ January-March 2021

71.83 ± 23.2 at S1 to 85.07 ± 22.1 at S2 and to 90.65 ± 18.6 at S3 with \( P < 0.0001 \) [Table 1]. A total of 1440 distractors were assessed. The number of MCQ items with NFDs reduced from 107 (52.20%) at S1 to 65 (31.7%) at S2 to 33 (16.1%) at S3; \( P < 0.0001 \) [Table 2]. The frequency of MCQ items with abnormal IOD reduced from 91 (53.53%) at S1 to 54 (31.76%) at S2 to 25 (14.71%) at S3; \( P < 0.0001 \) [Table 2]. The frequency of MCQ items with abnormal DI reduced from 30 (63.83%) at S1 to 12 (25.53%) at S2, and 5 (10.64%) at S3; \( P < 0.0001 \) [Table 2]. Figure 1 displays the trend of some of the indicators from the IA parameters over the 3-year period.

Table 1: The results of comparison of the item analysis parameters (difficulty index, distractor efficiency, and index of discrimination) across the three sessions

| Course | Parameter              | ANOVA P |
|--------|------------------------|---------|
| MEV    | DI                     | 0.0001  |
|        | DE                     | 0.001   |
|        | IOD                    | 0.0001  |
| CNS    | DI                     | 0.0006  |
|        | DE                     | 0.0054  |
|        | IOD                    | 0.0030  |
| MED    | DI                     | 0.9766  |
|        | DE                     | 0.1318  |
|        | IOD                    | 0.0004  |

| Course | Parameter              | Frequency (%) |
|--------|------------------------|---------------|
| MEV    | NFDs                   | 39 (56.41)    |
|        | Negative PB            | 19 (52.63)    |
|        | Abnormal index of discrimination | 37 (54.05) |
|        | Abnormal difficulty index | 16 (52.56)   |
| CNS    | NFDs                   | 78 (52.56)    |
|        | Negative PB            | 19 (73.06)    |
|        | Abnormal index of discrimination | 61 (68.85) |
|        | Abnormal difficulty index | 19 (73.68)   |
| MED    | NFDs                   | 88 (50.00)    |
|        | Negative PB            | 39 (48.72)    |
|        | Abnormal index of discrimination | 75 (42.67) |
|        | Abnormal difficulty index | 12 (58.33)  |
| All courses | NFDs              | 205 (52.20) |
|          | Negative PB          | 77 (55.84)   |
|          | Abnormal index of discrimination | 173 (53.53) |
|          | Abnormal difficulty index | 47 (63.83)  |

| MEV=Nan and his environment, MED=Internal medicine, CNS=Nervous system and special senses, DI=Difficulty index, DE=Distractor efficiency, CA=Cronbach’s alpha, ANOVA=Analysis of variance, SD=Standard deviation |

Table 2: The results of comparison of the item analysis parameters across the three sessions

| Course | Parameter              | Total | S1 | S2 | S3 | \( P \) |
|--------|------------------------|-------|----|----|----|--------|
| MEV    | NFDs                   | 39    | 22 | 11 | 6  | 0.0001 |
|        | Negative PB            | 19    | 10 | 7  | 2  | 0.015  |
|        | Abnormal index of discrimination | 37 | 20 | 13 | 4  | 0.0001 |
|        | Abnormal difficulty index | 16  | 9  | 5  | 3  | 0.02   |
| CNS    | NFDs                   | 78    | 41 | 25 | 12 | 0.0001 |
|        | Negative PB            | 19    | 14 | 5  | 0  | 0.001  |
|        | Abnormal index of discrimination | 61 | 42 | 12 | 7  | 0.0001 |
|        | Abnormal difficulty index | 19  | 14 | 4  | 1  | 0.002  |
| MED    | NFDs                   | 88    | 44 | 29 | 15 | 0.0001 |
|        | Negative PB            | 39    | 19 | 15 | 5  | 0.006  |
|        | Abnormal index of discrimination | 75  | 29 | 29 | 14 | 0.002  |
|        | Abnormal difficulty index | 12  | 7  | 4  | 1  | 0.09   |
| All courses | NFDs              | 205   | 107| 65 | 33 | 0.0001 |
|          | Negative PB          | 77    | 43 | 27 | 07 | 0.0001 |
|          | Abnormal index of discrimination | 173 | 91 | 54 | 25 | 0.0001 |
|          | Abnormal difficulty index | 47  | 30 | 12 | 10 | 0.0001 |

NFDs=Nonfunctional distractors, PB=Point biserial, MEV=Man and his environment, MED=Internal medicine, CNS=Nervous system and special senses

Table 3: The distribution of the nonfunctional distractors across the three sessions

| Number of NFDs | S1 | S2 | S3 | Total |
|---------------|----|----|----|-------|
| 1 distractor  | 80 | 54 | 25 | 159   |
| 2 distractors | 25 | 10 | 7  | 42    |
| 3 distractors | 2  | 2  | 0  | 4     |
| Total         | 108| 66 | 34 | 205   |

NFDs=Nonfunctional distractors

from 107 (52.20%) at S1 to 65 (31.7%) at S2 to 33 (16.1%) at S3; \( P < 0.0001 \) [Table 2]. Table 3 shows the distribution of NFD over the three sessions. Similarly, the number of MCQ items with NFDs pursued a downward trend from 43 (55.84%) at S1, 27 (35.06) at S2, and 07 (9.09%); \( P < 0.0001 \) [Table 2]. The number of MCQ items with abnormal IOD reduced from 91 (53.53%) at S1 to 54 (31.76%) at S2 to 25 (14.71%) at S3; \( P < 0.0001 \) [Table 2]. The frequency of MCQ items with abnormal DI reduced from 30 (63.83%) at S1 to 12 (25.53%) at S2, and 5 (10.64%) at S3; \( P < 0.0001 \) [Table 2]. Figure 1 displays the trend of some of the indicators from the IA parameters over the 3-year period.
The trend of item analysis parameters over the three sessions. S1 = First session, S2 = Second session, S3 = Third session. Adi = Number of abnormal/outside normal range difficulty index, npb = Number of negative point biserial, aiod = Number of abnormal/outside normal range index of discrimination, nfd = Number of nonfunctioning distractor

The mean CA for the examinations in S1, S2, and S3, respectively, was $0.51 \pm 0.1, 0.77 \pm 0.09$, and $0.84 \pm 0.03$, $P < 0.0001$ [Table 1]. Similar results were obtained when the same analysis was conducted on the individual courses [Tables 1 and 2].

DISCUSSION

Overall, the current study showed significant and graded upward trend in the quality of MCQ items constructed by the faculty members of UBCOM following repeated FDP over a 3-year period in the medical school. Our finding is consistent with reports from studies conducted elsewhere.\[13,18\] Although a number of studies had evaluated the effect of FDP on skills and behavior of medical educators, most of these studies carried out their intervention immediately after short-term interventional faculty development activities.\[19-21\] Conversely, the current study evaluated the influence of repeated FDPs offered over a 3-year period on item-writing skills. A well-constructed MCQ is an effective tool to evaluate varying levels of cognition ranging from comprehension, application, analysis, to synthesis among students.\[22\] In order to standardize MCQs, effort should be made toward frequent evaluation of the question through item analysis in order to make a reliable pool of MCQs.\[23\]

Our study showed graded improvement in the DE and IOD of MCQ items over the 3-year period of the study. The item discrimination index which is a measure of how well an item can distinguish between examinees who are knowledgeable and those who are not, or between masters and nonmasters, is dependent on whether a question item has desired discriminating capacity or not. This attribute of a good MCQ is partly a reflection of how the item in question is constructed. Thus, the improvement seen in the current study mirrors an improved question item-writing capacity, over time, on the part of the tutors that constructed the items.

Inherent in the concept of IA is its ability to improve items which will be used again in later tests.\[24\] In psychometric analysis, a crucial element in the quality of an MCQ item is the quality of the item’s distractors. None of the ID, IOD, or CA is sufficiently capable of assessing the performance of the incorrect response options and hence the importance of a distractor analysis. A distractor analysis, a response distribution analysis, addresses the performance of the incorrect response options. Our study showed that with repeated faculty development education activity, there is a gradual reduction in the number of nonfunctioning distractors which are MCQ options or alternatives that are selected by examinees in < 5% of cases.\[16\]

One area where many MCQs may fall below standard is in having effective and functioning distractors. Like in other forms of education, tutors in medical education frequently spend so much time on constructions of question item stems and much less time on developing plausible alternatives to the key answer. As the quality of MCQ items is partly hinged on these alternatives, the construction of distractors becomes as important as the key answers, therefore, the same measure of effort should be put into the construction of the wrong options.\[25\] Nevertheless, it is hard to underestimate the fact that crafting of MCQ item stem and key may not be a simple process and that searching for or developing plausible distractors may be very intricate.

Given the import of distractors as a tool in the hands of tutors to redirect students’ learning for the betterment, adjustments can, and should, be made to the options that are nonfunctioning. They are often replaced with potentially functional options or removed from the item.\[3,16,25\] This approach is well illustrated by Haladyina and Downing,\[7\] by studying functioning distractors in about 477 items on four MCQ assessments. They reported that over 38% of NFD were removed from the test questions they reviewed.

In modern education, one major concern in the writing of MCQ test for any examination is the reliability of the test items and test scores as a whole. To this end, PB analysis and CA analysis have proven to be suitable statistical tools. Aside from being a measure of reliability, CA also assesses the internal consistency of the test as a whole. Our study showed that repeated FDP upgraded the reliability as well as internal consistency of tests designed by faculty members.

Given that teachers’ knowledge in assessment and evaluation is not a static process but rather a complex, dynamic, and ongoing activity,\[26\] the need for education and re-education of medical educators cannot be overemphasized. The utility of assessments goes beyond the understanding of the student’s prowess at the learning outcome, be it cognitive, affective, or psychomotor. It includes the enhancement of modifications in teaching process that is hinged on students needs, alignment with specific learning outcome, coverage of specified course resources, and developing a more effective and comfortable learning environment.\[24,27\] Accordingly, the medical and
nonmedical educators alike are expected to contemporize their knowledge regarding MCQ item writing and assessment practices in general. It is critical for teachers to be well versed in the act of MCQ item writing to enable them reliably evaluate student learning achievement. FDP, as shown in the current study, is an easy way of achieving this goal.

**Limitations of the study**

In spite of the applicable outcome of our study, the study is not without some limitations. First, the study was premised on the assumption that all the faculty members involved in writing of the question items sourced their knowledge and understanding of MCQ item writing from the weekly FDP in the college. Second, only three out of the several courses were selected for the study. Third, a case–control design would probably have been more suitable but for the ethical consideration of such study. Nonetheless, we are of the opinion that FDP is also supposed to gear the faculty members toward SDL with the view to improving themselves. We also believed that the use of random selection technique in arriving at the selected courses makes the outcome of the current study reliable.

**Strength of the study**

Unlike many other studies on the subject matter that assessed a snapshot impact of FDP, the current study explores improvement in item-writing ability by following up MCQ item-developing ability of the same cohort of tutors over a period of 3 years, hence a better reflection of what obtains in reality. Collectively, this study is an outcome-focused effort representing a desirable progress toward demonstrating the influence of FDPs on the faculty and their question-developing achievements.

**Conclusion**

This study showed that repeated FDP improves the quality of MCQ items developed by faculty members. Therefore, to foster change and seed the much-needed reform in the training of medical students, the need for active training and re Training of medical educators on assessment techniques cannot be overemphasized.

**Acknowledgment**

The authors would like to acknowledge the staff of UBCOM, Saudi Arabia.

**Financial support and sponsorship**

Nil.

**Conflicts of interest**

There are no conflicts of interest.

**References**

1. Drew S. Perceptions of what helps learn and develop in education. Teach High Educ 2001;6:309-31.
2. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990;65:S63-7.
3. 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.
4. 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.
5. Tarrant M, Ware J. A framework for improving the quality of multiple-choice assessments. Nurse Educ 2012;37:98-104.
6. Schuworth LW, van der Vleuten CP. Different written assessment methods: What can be said about their strengths and weaknesses? Med Educ 2004;38:974-9.
7. Downing SM. The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. Adv Health Sci Educ Theory Pract 2005;10:133-43.
8. Vyas R, Supe A. Multiple choice questions: A literature review on the optimal number of options. Natl Med J India 2008;21:130-3.
9. Abdalla ME. What does item analysis tell us? Factors affecting the reliability of MCQs. Gezira J Health Sci 2011;7:17-25. Available from: https://www.academia.edu/2978805/What_does_item_analysis_tell_us_Factors_affecting_the_reliability_of_MCQs. [Last accessed on 2020 Feb 14].
10. Tavakol M, Dennick R. Post-examination analysis of objective tests. Med Teach 2011;33:447-58.
11. Harden RM, Crosby J. AMEE Guide No 20: The good teacher is more than a lecturer The twelve roles of the teacher. Med Teach 2000;22:334-47.
12. Ibrahim M, Al-Shahrani A. Implementing of a problem-based learning strategy in a Saudi medical school: Requisites and challenges. Int J Med Educ 2018;9:83-5.
13. Hartling L, Spooner C, Tjosvold L, Oswald A. Problem-based learning in pre-clinical medical education: 22 years of outcome research. Med Teach 2010;32:28-35.
14. Anderson LW. Curricular alignment: A Re-examination. Theory Pract 2002;41:255-60.
15. 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.
16. 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.
17. Bland JM, Altman DG. Statistics notes: Cronbach’s alpha. BMJ 1997;314:572.
18. Armstrong EG, Doyle J, Bennett NL. Transformative professional development of physicians as educators: Assessment of a model. Acad Med 2003;78:702-8.
19. Cole KA, Barker LR, Kolodner K, Williamson P, Wright SM, Kern DE. Faculty development in teaching skills: An intensive longitudinal model. Acad Med 2004;79:469-80.
20. Elliot D, Skeff KM, Stratos GA. How do you get to the improvement of teaching? A longitudinal faculty development program for medical educators. Teach Learn Med 1999;11:52-7.
21. Gruppen LD, Frohna AZ, Anderson RM, Lowe KD. Faculty development for educational leadership and scholarship. Acad Med 2003;78:137-41.
22. Masters K. Designing and Managing Multiple Choice Questions (2nd Ed). Available from: https://www.academia.edu/29261730/Designing_and_Managing_Multiple_Choice_Questions_2nd_Ed_. [Last accessed on 2020 Feb 14].
23. Dionaldo ML, D'Sa JL. Analysis of multiple choice questions: Item difficulty, discrimination index and distractor efficiency. Int J of Nursing Educ 2012;62:142-7.
24. Elliot D, Skeff KM, Stratos GA. How do you get to the improvement of teaching? A longitudinal faculty development program for medical educators. Teach Learn Med 1999;11:52-7.
25. Masters K. Designing and Managing Multiple Choice Questions (2nd Ed). Available from: https://www.academia.edu/29261730/Designing_and_Managing_Multiple_Choice_Questions_2nd_Ed_. [Last accessed on 2020 Feb 14].
26. Dionaldo ML, D’Sa JL. Analysis of multiple choice questions: Item difficulty, discrimination index and distractor efficiency. Int J of Nursing Education 2017;9:110-4. Available from: https://www.academia.edu/34857455/Analysis_of_Multiple_Choice_Questions_Item_Difficulty_Discrimination_Index_and_Distractor_Efficiency_ [Last accessed on 2020 Feb 14].
27. Pojham WJ. Classroom Assessment: What Teachers Need to Know. 2nd ed. Allyn & Bacon a Viacom Company, 160 Gould St; 1999.