

ORIGINAL ARTICLE

Evaluating the Effectiveness of a Faculty Development Program in Enhancing the Quality of Multiple-Choice Questions: A Quantitative Study

Saba Amjad¹, Noor Ijaz², Zakia Saleem¹, Kinza Aslam¹, Nabila Kaukab¹, Hajera Ahmed¹



Affiliation:

¹ University College of Medicine & Dentistry, University Of Lahore, Pakistan

Correspondence:

Dr. Zakia Saleem, Assistant Professor, Department of , University College of Medicine & Dentistry, University of Lahore, Pakistan

E-mail: zakiasaleem03@gmail.com

Received: Nov 30, 2025, **Accepted:** June 22, 2026, **Published:** June 30, 2026

DOI: <https://doi.org/10.47489/szmc.v40i2.876>

Data Availability: The corresponding author can provide access upon request

Funding: No funding received

Competing Interest: No competing interest to declare

Author Contributions:

SA, NI, ZS, KA, NK, HA :

- Each author made substantial contributions to the conception and design of the study, or acquisition, analysis, and interpretation of data.
- All authors were involved in drafting the manuscript or critically revising it for important intellectual content.
- All authors approved the final version of the manuscript to be published and agree to be accountable for all aspects of the work.

ABSTRACT

Background: Multiple choice questions (MCQs) are widely used in medical education for their objectivity and efficiency. However, poorly constructed questions can compromise the validity and reliability of the assessments. Faculty development programs have emerged as a key strategy to improve the quality of MCQs.

Objective: This study aims to evaluate the impact of structured FDPs on the MCQs construction using objective metrics and participants' feedback.

Methods: This observational study was conducted over three years period (2022 - 2024) at University College of Medicine & Dentistry. Three workshops on MCQs development were organized for the faculty. Feedback was collected through structured forms using 5- point Likert scale. Additionally, MCQs submitted before and after the workshops were analyzed for item flaws, difficulty & discrimination indices. The study also evaluated the impact of training using the Kirkpatrick Model. The data was computed and analyzed using SPSS version 26.0. The Shapiro Wilk test determined the normality of the data. P value <0.05 was considered as significant.

Result: A significant reduction in test-wiseness flaws (logical cues from 12.9% to 1.3%, $p < 0.001$) and irrelevant difficulty (extraneous details from 7.6% to 0.9%, $p < 0.001$) was observed. The proportion of moderately difficult items increased from 56.9% to 62.2%, while the proportion of very good discriminating items increased from 2.2% to 6.7%; however, neither change reached statistical significance ($p = 0.062$ and $p = 0.098$, respectively). Feedback from the participants revealed increasing satisfaction across workshops, with the highest ratings in workshop 3 (2024). Instructor expertise remained consistently rated above 87.5%.

Conclusion: Incorporation of multiple structured faculty development workshops lead to measurable improvements in MCQ quality. These training sessions can enhance the reliability and validity of medical assessments, benefiting both educators and learners.

Keywords: Faculty, Multiple choice questions, Program evaluation, Faculty Development, Assessment, Psychometrics

INTRODUCTION: Multiple-choice questions (MCQs) are widely used in medical education assessments due to their objectivity, reliability and efficiency in evaluating students' cognitive skills. They are considered the gold standard for written assessments. However, the effectiveness of MCQs depends heavily on their construction. Poorly designed MCQs can lead to misleading assessments, failing to measure students' critical thinking and knowledge application accurately.

Research has shown that a significant proportion of MCQs contain structural flaws, including poorly framed stems, implausible distractors and grammatical cues, which compromise their validity [1,2]. Faculty development programs (FDPs) have been implemented globally to address this issue & improve item writing skills. These training programs have demonstrated positive outcomes in terms of item quality and alignment with assessment standards [3].

Despite the recognized importance of FDPs, limited research has systematically evaluated the impact of multiple, structured FDPs on MCQ quality, particularly using standardized metrics like difficulty &

discrimination indices, distractor efficiency and item flaw analysis [4,5]. Many institutions rely on faculty experience rather than structured training [6] resulting in inconsistent question quality.

The present study addresses this gap by comparing two MCQ pools before and after successive FDP sessions, targeting the same cognitive level and course content. It aims to assess the impact of structured FDP on the quality of MCQ and provide evidence to support policy-level integration of such trainings within institutional assessment frameworks. To evaluate the effectiveness of faculty development programs in improving the quality of Multiple-Choice Questions (MCQs) used in high-stake assessments.

METHODS: It is an observational analytical study conducted in University College of Medicine & Dentistry, Lahore for 3 years from January 2022 to January 2024. A total of 450 MCQs including 225 MCqs from CBA 1 (2022) and 225 MCQs from CBA II (2024) were analyzed. The Ethical approval was obtained from the IRB of UCMD, UOL (Ethical Approval No. UOL/IREB/25/15/06/94 dated 11-12,-25). A total of 85 faculty members (30, 25 and 30 participants respectively) attended three MCQs development workshops conducted during the study period were included. Each workshop had an average of 28 participants. The same faculty members could not be followed longitudinally across all workshops due to faculty turnover. However, the participants involved in MCQ formation were from comparable disciplines and academic ranks. The non-Probability Convenience Sampling technique was used.

Inclusion Criteria:

Faculty members responsible for creating and reviewing MCQs for undergraduate assessments were included. Those who actively participate in the faculty development workshop and individuals contributing pre- and post-workshop MCQ pools for evaluation were included in the study.

Exclusion Criteria:

Faculty members who do not contribute MCQs for assessment were excluded from the study. Those who did not complete the faculty training workshop were also excluded.

Participation in faculty development workshops and feedback surveys was voluntary. All MCQ records pre and post workshop submission were anonymized prior to analysis and confidentiality was maintained throughout the study. The informed consent was taken from the assessment cell for the use of their MCQ bank required for this study.

After receiving the ethical approval from the Ethical Review Board (ERB), MCQs from the Combined Block Assessment (CBA 1) conducted in 2022 and 2024 were collected from the assessment cell. Details of the workshops held in 2023 & 2024 were obtained from the Faculty Development Program Committee. All submitted MCQ's were reviewed to assess the improvement in MCQ's quality before and after the workshop. The MCQ's were analyzed based on the item flaw, difficulty and discrimination indices [7]. Difficulty and discrimination indices were calculated as continuous measures and later categorized based on standard assessment criteria. As the frequency distribution of MCQs before and after the workshop were compared by inferential statistics, the Chi-square test was applied. MCQ defects were measured from the item-writing guidelines given by National Board of Medical Examiners (NBME). A standardized checklist was modified from Haladyna et al.'s principles of MCQ building. Each MCQ was reviewed independently by at least two trained medical education specialists having a minimum experience of at least 5 years in the assessment department and item writing. The reviewers were blinded to whether the MCQs were taken from pre-workshop or post-workshop pool. Discussions and consensus were used to resolve any discrepancies. The inter-rater reliability was not formally assessed in the study and is acknowledged as limitation of the study. The data was computed and analyzed using SPSS version 26.0. The Shapiro Wilk test determined the normality of the data. P value <0.05 was considered as significant. Feedback from the workshop participants was collected through structured Google forms, using 5-point Likert scale ratings (1=Poor, 5=Excellent) to rate content clarity, usefulness, organization, instructor knowledge, presentation style & responsiveness. The medical education experts reviewed the questionnaire and minor corrections were incorporated to ensure content and face validity, prior to data collection. The feedback played a crucial role in identifying and correcting flaws, ensuring improved question quality prior to next time submission. This method of evaluation of effectiveness of training programs is based on Kirkpatrick's model of evaluation [8].

Workshop details:

Workshop 1 (January 2023): Included 30 participants from both basic and clinical sciences.

Workshop 2 (June 2023): Attended by 25 faculty members from various departments.

Workshop 3 (April 2024): Conducted as a hand-on training session with 30 participants. Participants required prior registration and use of laptop for item construction exercises. Feedback forms were filled by 23, 8, 24 participants in workshops 1,2,3 respectively with response rates of 76.7%, 32% and 80% respectively

The categorical variables in the given study were summarized by the percentages and frequencies. For the comparison of item-writing flaws, Chi-square or Fisher's Exact tests were applied; as appropriate. Overall distributions for difficulty and discrimination index, Chi-square tests were used for analysis and comparison. The confidence intervals of ninety five percent (95% CI) were considered where applicable. As categories having multiple item-flaws were evaluated, results were interpreted with caution so as to minimize any type of false positive errors.

Analysis of item flaws and indices was conducted using standardized procedures to assess improvement in MCQ quality pre- & post-training.

RESULTS: The participants of the workshops were representing various departments including Basic & Clinical sciences departments. The faculty members belong to various professional ranks including Professors, Associate Professors, Assistant Professors, Senior Registrars, registrars, senior demonstrators, demonstrators, and postgraduate residents, representing a wide range of experience levels. Most of the participants had not attended the workshop on MCQ's development before indicating their keen interest in improving MCQ's development skills.

1. Issues Related to Test-Wiseness

A comparative analysis of test-wiseness-related flaws in MCQs revealed significant improvement after the faculty development workshops. In CBA1 conducted in 2022, logical cues were present in 29 MCQs (12.9%), which significantly reduced to 3 MCQs (1.3%) in CBA1 conducted in 2024 ($p < 0.001$). Grammatical errors were also decreased from 9 (4.0%) in 2022 to 5 (2.2%) in 2024, though this reduction was not statistically significant ($p = 0.277$). Similarly, minor differences were observed in other categories such as absolute terms, word repeats, and convergence strategies, but these changes were not statistically significant (all $p > 0.05$) (Table 1). No cases of long correct answers were found in either pre and post workshop MCQs papers (Fig 1).

Table 1: Comparison of Issues Related to Test-Wiseness in MCQs Between CBA1-2022 and CBA1-2024

S. No.	Issues related to test wiseness	CBA1-22 (n=225)	CBA1-24 (n=225)	p-value #
1.	Grammatical error	9(4.0%)	5(2.2%)	0.277
2.	Logical Cues	29(12.9%)	3(1.3%)	<0.001*
3.	Absolute terms	4(1.8%)	1(0.4%)	0.177
4.	Long Correct Answers	0(0.0%)	0(0.0%)	N/A
5.	Word Repeats	4(1.8%)	3(1.3%)	>0.999
6.	Convergence Strategy	0(0.0%)	1(0.4%)	>0.999

#Chi-square test \ Fisher's exact test, *Significant

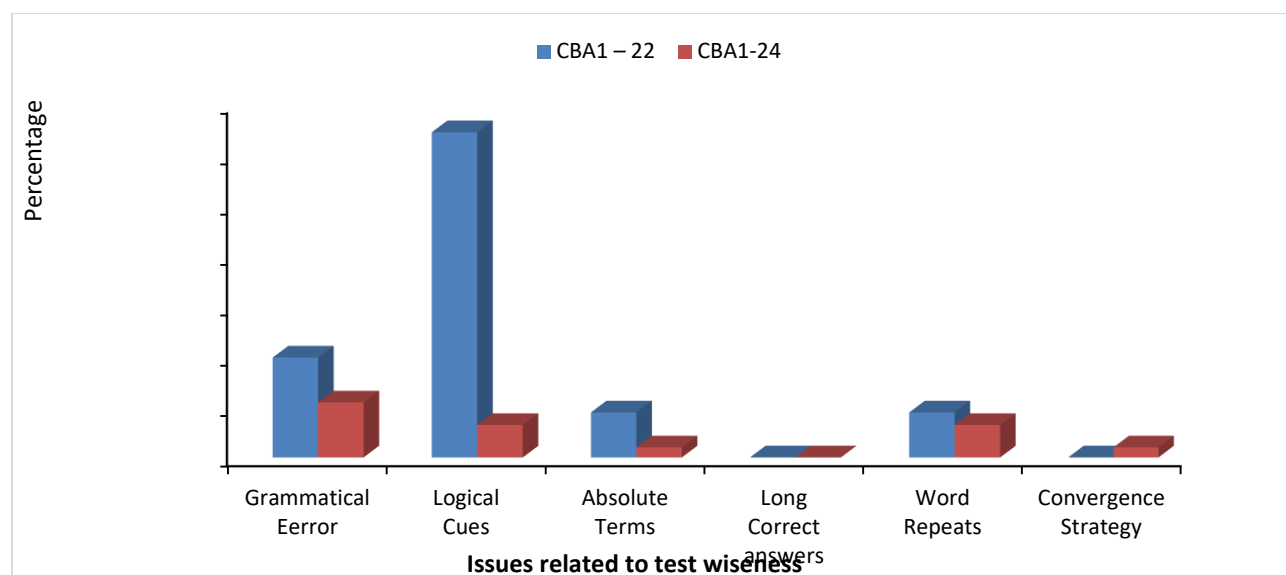


Fig 1: Bar chart indicating percentages of issues related to test-wisdom

There was a marked reduction in flaws associated with irrelevant difficulty in MCQs after the intervention. Extraneous details in stems dropped significantly from 17 (7.6%) in 2022 to 2 (0.9%) in 2024 ($p < 0.001$). The use of non-parallel language in options also showed a statistically significant reduction from 16 (7.1%) to 5 (2.2%) ($p = 0.014$). Although the presence of long, complicated or double options increased slightly (from 3.1% to 5.8%), this change was not statistically significant ($p = 0.170$). Other categories such as numeric inconsistency, vague frequency terms and tricky stems showed no significant differences (Table 2).

Table 2: Comparison of Issues Related to Irrelevant Difficulty in MCQs between CBA1-2022 and CBA1-2024

Issues Related to Irrelevant Difficulty	CBA-I 2022 (n = 225)	CBA-I 2024 (n = 225)	p-value#
Options are long, complicated, or double	7 (3.1%)	13 (5.8%)	0.170
Extraneous detail in stem	17 (7.6%)	2 (0.9%)	<0.001*
Numeric data not stated consistently	3 (1.3%)	2 (0.9%)	0.999
Vague frequency terms	0 (0.0%)	1 (0.4%)	0.999
Nonparallel language in options	16 (7.1%)	5 (2.2%)	0.014*
Stems are tricky or unnecessarily complicated	7 (3.1%)	9 (4.0%)	0.611
None of the above	0 (0.0%)	0 (0.0%)	N/A
All except	0 (0.0%)	0 (0.0%)	N/A

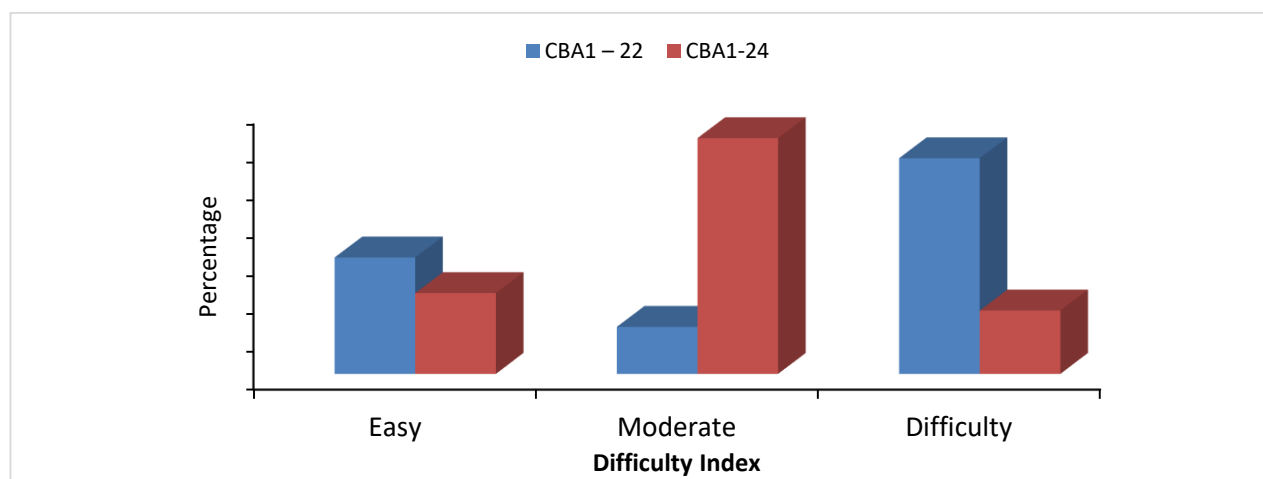
#Chi-square test \ Fisher's exact test, *Significant

Analysis of the difficulty index of MCQs showed a shift toward moderate-level items in 2024. Difficulty Index is measured as % of participants who got the item right. The value of Difficulty Index (easiness) ranges from 0-100% with percentage >78% is categorized as an easy item, 78-25% as acceptable & <25% as difficult. In CBA1-2022, 56.9% of the items were of moderate difficulty, which increased to 62.2% in 2024. The proportion of easy items decreased from 30.7% to 21.3%, while difficult items increased from 12.4% to 16.7% (Fig 2). A trend toward a greater proportion of moderately difficult items was observed; however, the difference did not reach statistical significance ($p = 0.062$) (Table 3).

Table 3: Comparison of the Difficulty Index in CBA Paper 1 Between 2022 and 2024

Difficulty Index	CBA-I 2022 (n = 225)	CBA-I 2024 (n = 225)	p-value#
Easy	69 (30.7%)	48 (21.3%)	0.062
Moderate	128 (56.9%)	140 (62.2%)	-----
Difficult	28 (12.4%)	37 (16.4%)	-----

p-value represents the overall Chi-square comparison of the difficulty-index distribution between groups

**Figure 2: Bar chart showing comparison of the difficulty index in CBA Paper 1 between 2022 and 2024**

The distribution of MCQs across discrimination index categories in CBA Paper 1 was compared between 2022 and 2024. Discrimination index is the ability of an item to discriminate between high and low achievers. Its value ranges from -1.0 to 1.0. negative value indicates non-discrimination while positive value shows discrimination [9]. The proportion of items classified as “Very Good” increased from 2.2% in 2022 to 6.7 % in 2024. The percentage of items rated as “Good” remained stable (14.2% to 15.6%), while those rated “Marginal” increased slightly from 28.9% to 31.1%.

A small decrease was observed in “Poor” items (44.0% to 40.0%) and in “Very Poor” items (10.7% to 6.7 %). Although these trends suggest gradual improvement in item quality, particularly a reduction in very poor items and an increase in very good items, however, this difference was not statistically significant ($p = 0.098$) (Table 4).

Table 4: Comparison of the discrimination index in CBA Paper 1 between 2022 and 2024

Discrimination Index	CBA1-22 (n=225)	CBA1-24 (n=225)	p-value#
Very Good	5 (2.2%)	15(6.7%)	0.098
Good	32(14.2%)	35(15.6%)	
Marginal	65(28.9%)	70(31.1%)	
Poor	99(44.0%)	90(40.0%)	
Very Poor	24(10.7%)	15(6.7%)	

p-value represents the overall Chi-square comparison of discrimination-index distribution between groups.

The feedback analysis across the three MCQ development workshops reveals distinct patterns in participant satisfaction and areas for growth, with workshop 3 (2024) receiving the highest ratings (75-87.5% positive responses on key metrics), suggesting improvements in content delivery and structure, while workshop 2 (2023) demonstrated polarized feedback (50% positive ratings) that may be attributed to its smaller participant cohort (n=8) or session design factors, though all workshops maintained consistently exceptional ratings (>85% positive) for instructor performance, highlighting faculty expertise as a universal strength, with qualitative feedback emphasizing a recurring demand for more interactive hands-on activities and advanced MCQ design techniques in future iterations to further enhance learning outcomes and practical application.

Table 5: Descriptive Summary of Participant Evaluation Scores Across Faculty Development Workshops

Evaluation Criteria	Workshop1 (2023)	Workshop2 (2023)	Workshop3 (2024)
Objectives clearly explained	60.9%(4-5)	50.0%(4-5)	75.0%(4-5)
Useful material	69.6%(4-5)	50.0%(4-5)	81.3% (4-5)
Instructor knowledge	100%(4-5)	87.5%(4-5)	87.5%(4-5)
	100%	50.0%	87.5%
Overall workshop rating	(Good/Excellent)	(Good/Excellent)	(Good/Excellent)

* Percentages represent participants assigning ratings of 4 (Very Good) or 5 (Excellent) on the five-point Likert scale.

DISCUSSION: The multiple-choice question is one of the widely used assessment tool across all levels of education. It is essential for the faculty members to properly construct the MCQ's. However, most faculty members are either not familiar with standard MCQ making guidelines or reluctant to change their practices [3]. The findings of the current study showed that the multiple faculty training workshops contribute to measurable improvement in item quality, particularly by reducing flaws related to test-wiseness and irrelevant difficulty. A statistically significant decrease in extraneous detail in stem (7.6% to 0.9%), logical cues (12.9% to 1.3 %) and non-parallel language in options (7.1% to 2.2%) after various FDP. The current study depicted the positive influence of workshops in minimizing the flaws of test-wiseness and irrelevant difficulty. The shift towards moderate difficulty in post-item analysis reflects a better alignment with assessment standards, aiming to evaluate students critical thinking. A statistically significant decrease in the use of logical cues was observed in CBA 1 from 12.9% in 2022 to 1.3% in the year 2024 ($p < 0.001$) Table 1. Logical cues indicate the most important construct flaws, enabling students to choose correct answers without proper understanding. This decline in the logical cues indicated the increased awareness among faculty members. This awareness helped the faculty to avoid inadvertent assistance to the students through poor stem phrasing or distractors. However other flaws like grammatical errors, word repeats, convergence strategies and absolute terms also showed a decline but the changes were not statistically significant. These findings are consistent with the results of the study which reported improvement in the skills of MCQ construction and decrease in item flaws [1]. Another study conducted in 2021, also indicated similar findings by using Kirkpatrick's model. It proposed that MCQ development workshops can significantly enhance the ability of faculty members to identify flaws and construct high quality MCQ's. The results of the MCQ's after the series of MCQ development workshop demonstrated significant reduction in irrelevant difficulty flaws [8]. A marked improvement was observed in the reduction of extraneous details in MCQ stem, decreasing from 7.6% to 0.9% ($p \text{ value} < 0.001$) along with a noticeable reduction in non-parallel language in options from 7.1% to 2.2% ($p \text{ value} = 0.014$) Table 2. These findings are consistent with the findings provided by NBME which focus on clarity and uniformity in item structure to minimize items with

irrelevant difficulty that affects the validity and reliability. Study by Rauf and Sultans [10], concluded the similar findings that incorporation of workshops leads to a significant reduction in item flaws, enhancing cognitive level and fair evaluations. Furthermore, the proportion of long option ranging from 3.1% to 5.8% was not statistically significant ($p=0.170$) (Table 2) in year 2022 to 2024 respectively indicating that there is a need of contextualization. This is required for evaluating critical thinking and to avoid irrelevant difficulty. These findings are in correspondence with level 3 (behavior) of Kirkpatrick's evaluation model which suggested that the participants have applied the principles of item writing in practice, effectively.

This study revealed a positive drift in the MCQ quality after the implementation of structured faculty development program. The difficulty index of CBA paper depicted a shift towards a more balanced assessment with an increased proportion of moderately difficult items from 56.9% in 2022 to 62.2% in CBA paper 2024. Concurrently, a decrease in easy item from 30.7% to 21.3%. and an increase in number of difficult items was found from 12.4% to 16.7% in year 2022 and 2024, respectively (Table 3). Although these changes were not significant ($p=0.062$) but these trends showed a progress towards an optimal difficulty distribution.

Similarly, the discrimination index analysis showed an increase in improvement in quality of items, with very good items increasing from 2.2% to 6.7% and decrease of very poor items from 10.7% to 6.7%. However, the proportion of good items remained stable with a minor upward shift in marginal items 28.9% to 31.1% and a decline in poor items (44.0% to 40.0%). Although a higher proportion of very good items and fewer very poor items were observed in 2024, the overall change in discrimination-index distribution was not statistically significant ($p=0.098$). (Table 4)

While the workshops successfully addressed MCQ design principles, the feedback highlights opportunities to expand into more nuanced areas, such as item analysis and difficulty calibration. The quantitative analysis across three workshops showed a positive trend of progressive improvement. The workshop 3 (2024) had the highest percentage (75%) of the participants agreed that the objectives of the workshop were clearly defined, 81.3% rating the workshop material was useful and 87.5% expressing satisfaction. While workshop 2 showed lower percentages across all categories (50%) potentially due to smaller participant number. The instructor's knowledge consistently had the highest percentage above 87.5% across all workshops. The findings of this study correspond to level I (Reaction) of the Kirkpatrick model, which assesses participants' immediate response to training programs. Such positive reactions are crucial, as they can influence the motivation to learn and apply new skills. This alignment with the Kirkpatrick Model is supported by recent studies demonstrating high participant satisfaction and significant knowledge gains, emphasizing the model's applicability in assessing educational interventions[11,12].

Despite these strengths, some metrics did not demonstrate statistically significant improvement. This may be attributed to the small sample size, variable level of faculty engagement or the need of more frequent sessions. Nonetheless, these findings are consistent with the literature indicating that iterative, multi-session FDPs have greater, long-term impact than single sessions[13,14].

To maintain the progress, the institution should implement regular pre and post assessment item reviews and peer feedback mechanisms. Further studies should also explore additional variables such as distractor efficiency and conduct qualitative evaluations of faculty perceptions to enrich understanding.

CONCLUSION: This study concludes that structured faculty development workshops improve the quality of MCQ's used in undergraduate medical assessments. These workshops were associated with reductions in certain item-writing flaws, mainly logical cues, unnecessary details and inconsistent language. Although improvements were observed in difficulty and discrimination indices, but these results are not statistically significant. Participants reported satisfaction with workshop structure, content and facilitation. Overall, the findings suggest that integration of regular, structured FDPs serve as a major component of institutional quality assurance strategies in medical education.

It is recommended that to improve the construct of multiple-choice questions, there should be implementation of regular and structured faculty development training programs focusing mainly on item flaws and item analysis. Routine post hoc analysis of items would improve the quality of assessment and provide timely constructive feedback to faculty members.

Due to the faculty turnover, the same group of participants could not be followed across the three workshops. In addition, faculty perceptions were not explored through interviews or focus groups, which may have enriched the findings.

REFERENCES

1. Beg MA, Tabassum T, Ali S. Role of faculty development workshop for improving MCQS quality in basic medical sciences. *Biomedica* – Vol. 37, Issue 1, 2021. <http://doi.org/10.51441/BioMedica/5170>
2. Ansari M, Sadaf R, Akbar A, Rehman S, Chaudhry ZR, Shakir S. Assessment of distractor efficiency of MCQS in item analysis. *The Professional Medical Journal*. 2022 Apr 30;29(05):730-4. <https://doi.org/10.29309/TPMJ/2022.29.05.6955>
3. Agarwal A, Khattak O, Esmaeel SE, Fahmy E, Bayomy NA, Mehmood SI, Mohamed H, Bawadekji A, Alotibi FA, Alanazi MS, Younes A. Efficacy of Faculty Development Training Workshops (FDTWs) on Writing High-Quality Multiple-Choice Questions at Northern Border University (NBU) in the Kingdom of Saudi Arabia (KSA). *Cureus*. 2024 Jun 18;16(6). DOI: [10.7759/cureus.62607](https://doi.org/10.7759/cureus.62607)
4. Mashood S. Quality Assurance of Multiple-Choice Questions Test Through Item Analysis. *Life and Science*. 2023 Oct 5;4(4):7-Life & Science 2023 Vol. 4, No. 4. Doi: <http://doi.org/10.37185/LnS.1.1.315>
5. Baste VS. Item analysis of MCQs in physiology and its correlation with faculty's perception of difficulty level of MCQs. *National Journal of Physiology, Pharmacy and Pharmacology*. 2023 Jun 30;13(7):1444-. DOI: [10.5455/njppp.2023.13.12601202220122022](https://doi.org/10.5455/njppp.2023.13.12601202220122022)
6. Kiran F, Ayub R, Rauf A, Qamar K. Evaluating the impact of faculty development programme initiative: are we really improving skills in MCQ writing. *J Pak Med Assoc*. 2021 Oct 1;71(10):2434-8. DOI: <https://doi.org/10.47391/JPMA.1207>
7. Chauhan A, Khaliq F, Nayak KR. Assessing Quality of Scenario-Based Multiple-Choice Questions in Physiology: Faculty-Generated vs. ChatGPT-Generated Questions among Phase I Medical Students. *International Journal of Artificial Intelligence in Education*. 2025 Apr 7:1-30. <https://doi.org/10.1007/s40593-025-00471-z>
8. Ali R, Sultan AS, Zahid N. Evaluating the effectiveness of MCQ development workshop using cognitive model framework: A pre-post study. *JPMA. The Journal of the Pakistan Medical Association*. 2021;71(1 (A)):119. DOI: <https://doi.org/10.47391/JPMA.1068>
9. Rezigalla AA, Eleragi AM, Elhussein AB, Alfaifi J, ALGhamdi MA, Al Ameer AY, Yahia AI, Mohammed OA, Adam MI. Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*. 2024 Apr 24;24(1):445. <https://doi.org/10.1186/s12909-024-05433-y>
10. Rauf AR, Sultana S. Effect of faculty training on quality of Multiple Choice Questions. *Rawal Medical Journal*. 2021 Dec;46(2):430. Available at <https://www.rmj.org.pk/index.php?mno=133092>
11. Ahmad S, Imtiaz M, Aziz RS, Manzoor N, Ahmed S, Habib B. Kirkpatrick's Approach to Build Faculty Development Programs in Medical Education. *Pakistan Journal of Medical & Health Sciences*. 2023 Apr 29;17(03):441-. DOI: <https://doi.org/10.53350/pjmhs2023173441>
12. Shallal TM, Alkhateeb NE, Al-Dabbagh A. Virtual faculty development program in bioethics evaluated by Kirkpatrick model: A unique opportunity. *PloS one*. 2023 Oct 30;18(10):e0293008. <https://doi.org/10.1371/journal.pone.0293008>
13. Lee HY, Yune SJ, Lee SY, Im S, Kam BS. The impact of repeated item development training on the prediction of medical faculty members' item difficulty index. *BMC Medical Education*. 2024 May 30;24(1):599. <https://doi.org/10.1371/journal.pone.0293008>
14. Eleragi AM, Miskeen E, Hussein K, Rezigalla AA, Adam MI, Al-Faifi JA, Alhalafi A, Al Ameer AY, Mohammed OA. Evaluating the multiple-choice questions quality at the College of Medicine, University of Bisha, Saudi Arabia: a three-year experience. *BMC Medical Education*. 2025 Feb 13;25(1):233. <https://doi.org/10.1186/s12909-025-06700-2>

The Authors:

- Dr. Saba Amjad, Associate Professor, Department of Dentistry, University College of Medicine and Dentistry, The University of Lahore, Pakistan.
- Dr. Noor Ijaz, Assistant Professor, Department of Dentistry, Ameer Ud Din Medical College (Post Graduate Medical Institute), Lahore, Pakistan.
- Dr. Zakia Saleem, Assistant Professor, Department of Dentistry, University College of Medicine and Dentistry, The University of Lahore, Pakistan.
- Dr. Kinza Aslam, Associate Professor, Department of Dentistry, University College of Medicine and Dentistry, The University of Lahore, Pakistan.
- Dr. Nabila Kaukab, Professor, Vice Principal, Department of Dentistry, University College of Medicine and Dentistry, The University of Lahore, Pakistan.
- Dr. Hajera Ahmed, Senior Demonstrator, Department of Dentistry, University College of Medicine and Dentistry, The University of Lahore, Pakistan.