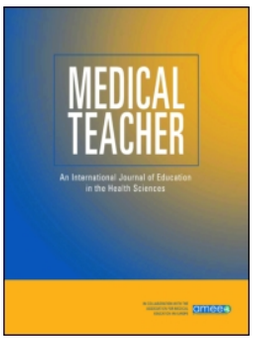

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: http://www.tandfonline.com/loi/imte20

Evaluating the long-term impact of faculty development programs on MCQ item analysis

Aliaa Amr Alamoudi, Basem Salama El-Deek, Yoon Soo Park, Lana Adey Al Shawwa & Ara Tekian

To cite this article: Aliaa Amr Alamoudi, Basem Salama El-Deek, Yoon Soo Park, Lana Adey Al Shawwa & Ara Tekian (2017): Evaluating the long-term impact of faculty development programs on MCQ item analysis, Medical Teacher, DOI: 10.1080/0142159X.2016.1254753

To link to this article: http://dx.doi.org/10.1080/0142159X.2016.1254753

Published online: 22 Jan 2017.

MEDICAL TEACHER, 2017
http://dx.doi.org/10.1080/0142159X.2016.1254753

MEDICAL TEACHER
Taylor & Francis
Taylor & Francis Group

Evaluating the long-term impact of faculty development programs on MCQ item analysis

Aliaa Amr Alamoudi[a], Basem Salama El-Deek[b], Yoon Soo Park[c], Lana Adey Al Shawwa[b] and Ara Tekian[c]

[a]Department of Clinical Biochemistry, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia; [b]Department of Medical Education, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia; [c]Department of Medical Education, University of Illinois - College of Medicine at Chicago, Chicago, IL, USA

ABSTRACT

Purpose: Evaluating the long-term impact of faculty development programs (FDPs) can help monitor the effectiveness of the program and identify areas for development. This study examined long-term differences in confidence, knowledge, behaviors, and policies of faculty members who attended FDPs on multiple choice question (MCQ) item analysis and faculty members who did not attend the FDPs.

Methods: A cross-sectional study design was used, by administering a 24-item survey to a representative sample (simple random selection) of 61 faculty members at King Abdulaziz University Faculty of Medicine.

Results: Among respondents, 34% did not attend FDPs; 53% attended 1–3 FDPs; and 13% attended more than 3 FDPs on MCQ item analysis. Results showed that faculty knowledge on elements of MCQ item analysis was significantly greater ($p = .01$) for members who attended the FDPs. Faculty who attended FDPs on MCQ item analysis were twice more likely to conduct item analysis in general ($p = .020$) and four times more likely to conduct item analysis for more than 70% of module examinations ($p = .005$).

Conclusion: FDPs focused on MCQ item analysis can yield systematic changes on faculty confidence, knowledge, and behaviors. Moreover, FDPs also need support from the department and need sustained strategic support to ensure continued effectiveness.

Introduction

Faculty development programs (FDPs) have become a crucial tool to support faculty members who are expected to succeed in multiple areas of academic medicine (Steinert et al. 2006; Leslie et al. 2013). At King Abdulaziz University Faculty of Medicine (KAU-FOM), the Medical Education Department was created to provide ongoing support and FDPs on knowledge, skills, and abilities to help faculty perform multiple duties. One of the main approaches for FDPs at KAU-FOM is to offer single training sessions and seminars to faculty members. A FDP commonly conducted is the assessment and psychometric review of item analysis that focus on multiple choice questions (henceforth, "FDPs on MCQ item analysis").

It is well known that multiple choice questions (MCQs) are widely used for the assessment of medical students and account for a large proportion of medical student examinations. Well-designed and high-quality MCQs are therefore essential to reflect students' performance in a course (Moss 2001; Hochlehnert et al. 2012). MCQ item analysis evaluates students' responses to test items (Moussa et al. 1991) and analyzing the quality and characteristics of individual items that ultimately contribute to the overall validity of a test score. As such, MCQ item analysis reports can facilitate the elimination of misleading or nonfunctioning items and allow the modification of items, which can be reused in future exams (Moussa et al. 1991).

Two commonly used item statistics are item difficulty and item discrimination indices. The item difficulty index can be represented as the percentage of students who correctly answered an item and can range from 0% to 100%. Higher levels indicate an easy item, while lower levels indicate item difficulty. The discrimination index, however, reflects the ability of an item to discriminate between low and high ability students. Items with high discrimination can discriminate the performance of students who scored high and those who scored low in an exam (Moussa et al. 1991). Assessments with higher discrimination indices have higher internal-consistency reliability.

FDPs on MCQ item analysis serve to deliver concepts of item analysis, aimed to provide participants with the knowledge and skills required to interpret item analysis reports and to use findings from the item analysis to modify MCQ

Practice points

- Faculty members who attended FDPs on MCQ analysis reported a long-term increase in confidence, knowledge, behaviors related to MCQ item analysis.
- Training sessions provided by FOM significantly increased the use and application of MCQ item analysis in actual testing contexts.
- A key factor for effective faculty development in MCQ item analysis is the strict implementation of policies.
- Clear objectives and the content of training sessions provided by FDPs should be distributed to the faculty before workshop implementation.

CONTACT Aliaa Amr Alamoudi aalamoudi2@kau.edu.sa Department of Clinical Biochemistry, Faculty of Medicine, King Abdulaziz University, P.O. Box 80200, Jeddah 21589, Saudi Arabia

items. Furthermore, these training sessions provide participants with insights on the role of MCQ item analysis in improving not only the quality of exams, but concurrently, the educational process as a whole. Despite the increasing scope of FDPs worldwide and its significant role, a persistent question remains to be the effectiveness of FDPs in inducing significant changes, as intended by their objectives (Naeem et al. 2012). It is therefore essential to conduct continuous and rigorous evaluation studies to measure outcomes of such programs. Evaluating educational outcomes of FDPs would have a significant role in improving their future design and implementation, thus ensuring their complete effectiveness (Leslie et al. 2013).

A commonly used conceptual framework for evaluating FDPs is the Kirkpatrick model (Kirkpatrick 1967). The model is based on four levels: learner's reflection, learning (knowledge and skills), behavioral changes, and results, which include changes in organizational practice. Despite significant improvements in recent years in FDP evaluations, a number of challenges remain. A systemic review conducted by Steinert on FDP studies emphasized the need for more longitudinal studies that can evaluate the effectiveness of expected changes due to FDPs over longer periods of time (Steinert et al. 2006). In addition, improvements in the methods of assessments and a more rigorous use of control groups wererecommended (Steinert et al. 2006). In a follow-up review by Leslie et al. (2013), a shift toward more longitudinal FDP evaluation studies were documented. However, participants' perceptions on the usefulness and satisfaction of the FDP were still measured immediately following the program, rather than tracking their long-term impact. In addition, knowledge and behavioral changes remained to be measured using subjective self-reported data. A better understanding of FDP impact and maintenance of change over time can help guide the development toward a constructive plan to follow-up on their programs (McAndrew et al. 2013)

Currently at KAU-FOM, evaluations of FDPs have generally been measured through immediate post-training surveys. In some occasions, participants have been assessed through pre- and post-tests measuring their knowledge, immediately following the training sessions. In this study, we aimed to evaluate the FDP, by measuring the long-term impact on confidence, knowledge, and change in behavior of participants who attended MCQ item analysis sessions, beyond their initial FDP participation. The objectives of the study were as follows: (1) to measure the percentage of KAU-FOM faculty who participated in MCQ item analysis training session provided by the Medical Education Department, (2) to determine differences in perceptions of confidence, knowledge, and behavior regarding MCQ item analysis between those who participated and those who did not, (3) to evaluate the level of knowledge through objective questions, and finally (4) to evaluate the changes in behavior among participants of FDPs on MCQ item analysis beyond their initial participation (intervention) to the training.

Methods

Curriculum

The curriculum at KAU FOM is divided between two phases: Phase I (preclinical) and Phase II (clinical, including four major clerkships). In Phase I, basic sciences are taught as core courses and system-based modules (Ayuob et al. 2015). Core course exams are generally developed, constructed, and administered by the department, while module exams are designed and administered by the interdisciplinary committee from various departments that teach within the module. As such, the impact of FDPs on differences in MCQ item analysis policies and faculty behaviors between department-based and interdisciplinary-based exams was examined in this study.

FDP MCQ item analysis sessions

FDP MCQ item analysis sessions provided by FOM are given either as part of an assessment workshop or as an individual session. All sessions are given by full- or part-time faculty members of the Medical Education Department. The main objectives of these sessions are for participants to define and identify elements of MCQ item analysis, with special emphasis on item difficulty and item discrimination indices, to interpret item analysis reports, and to value the importance of MCQ item analysis, as they contribute directly to the validity evidence of the assessment. Sessions are usually given as a three-hour session divided into theoretical and practical components, in which participants are exposed to MCQ item analysis reports to evaluate and discuss.

Study design

Participants

In a cross-sectional study, a representative sample of 61 KAU-FOM faculty members from various departments was selected randomly (simple random selection) and asked to complete a questionnaire evaluating their MCQ item analysis training sessions. The departments included Biochemistry, Anatomy, Physiology, Pharmacology, Microbiology and Parasitology, General Surgery, Internal Medicine, Pediatric, Obstetrics and Gynecology, Hematology, Orthopedics, and ENT. The inclusion criteria were faculty members who are Assistant Professors, Associate Professors, or Professors (full). Participants also included faculty who did not participate in MCQ item analysis training sessions provided by the FOM. Participants were requested not to refer to any previous material regarding MCQ item analysis while completing the survey (the survey included items on factual knowledge of item analysis).

Survey

A 24-item survey was developed and tested for initial validity evidence, including psychometric characteristics and item analysis. In order to ensure further validity (content and clarity of items in response process), the survey was tested by a focused group. The survey was then distributed to participants in hardcopy and administered, focusing on participants' perceptions of the FDP sessions on MCQ item analysis during the past 3 years. The survey was designed to cover the following topic areas: (1) attendance in FDP sessions on MCQ item analysis, (2) attitudes and beliefs (confidence) towards MCQ item analysis and training

sessions, (3) FOM policies on MCQ item analysis, (4) knowledge on MCQ item analysis, and (5) reflections on the practice of MCQ item analysis. To objectively assess respondents' knowledge of MCQs, items also included questions on item difficulty, item discrimination, and distractor options. FDP session materials on MCQ item analysis were provided by the Medical Education Department and session instructors.

Analysis

Data were collected from all 61 participants and statistical analyses were conducted using IBM SPSS 20 (Armonk, NY). Descriptive statistics were used to examine trends in data. Responses on the 24-item survey were cross-tabulated by frequency of FDPs attended on MCQ item analysis. Tests of proportions were conducted using the chi-squared test, both at the item level and also for the pooled responses by survey section. Ethical approval was obtained from the committee of ethics and research at KAU-FOM.

Results

Participant characteristics

Among the 61 participants, there were more junior faculty than senior faculty (Assistant Professor = 44%, Associate Professor =33%, Professor =23%). Nearly all participants responded to have knowledge on MCQ item analysis (92%). Over 52% attended 1–3 FDPs on MCQ item analysis within the last 3 years; 34% never attended FDPs on MCQ item analysis; and 13% attended more than 3 FDPs on MCQ item analysis within the last 3 years. Descriptive statistics of participants are summarized in Table 1.

Table 1. Descriptive statistics: %.

Factor	Percent (%)
Gender	
Male	44
Female	56
Academic position	
Assistant professor	44
Associate professor	33
Professor	23
Years of teaching	
0–5 years	18
5–10 years	12
>10 years	71
Knowledge on MCQ item analysis	
Yes	92
No	8
Attendance in FDP MCQ item analysis sessions by FOM	
None	34
1–3	53
3+	13

Confidence on benefits of MCQ item analysis

When evaluating faculty members' confidence on the impact of MCQ item analysis, 72% agreed that MCQ item analysis can improve the quality of exams and ultimately enhance students' learning. Overall, there was significant difference in the pooled proportions confidence with respect to number of FDPs attended (no FDP attendance, attendance on 1–3 FDPs, attendance on more than 3 FDPs), $p = .042$. However, there were no significant differences at the individual item levels (see Table 2).

Knowledge of MCQ item analysis

Faculty who participated in FDP sessions within the last 3 years had greater knowledge of item analysis than faculty who did not attend or had fewer attendance in the training sessions, $p = .010$. This is notable, given that 92% of all participants claimed knowledge on item analysis. In particular, there were significant differences in knowledge regarding the use of item difficulty, $p = .014$. Among faculty who did not attend the FDP sessions, less than 20% knew about the item difficulty range; moreover, only 5% knew about distractor options (see Table 2).

Behaviors toward item analysis

Overall, there were significant behavioral differences in faculty who attended FDP sessions within the last 3 years versus faculty who did not, $p < .001$ (see Table 2). At the item level, faculty who attended FDPs on MCQ item analysis were twice more likely to conduct item analysis in general ($p = .020$) and about four times more likely to conduct item analysis for more than 70% of module examinations ($p = .005$). Moreover, more than 75% of faculty who participated in FDPs reported to apply item analysis to modify exams, while this was only 48% among faculty who did not attend the FDPs.

Table 2. Survey response by number of FDP attendance on MCQ item analysis: column %.

		Number of FDPs attended (%)[b]				
Domain[a]	Question	None	1–3	3+	Total	*p* Value[c]
Confidence ($p = .042$)	1. Believe that MCQ item analysis training sessions are beneficial	67	69	100	72	.167
	2. Believe that MCQ item analysis improve student learning	81	78	100	82	.351
	3. Believe that MCQ item analysis improve exams	91	88	100	90	.568
Knowledge ($p = .010$)	1. Answered correctly on difficulty index range	19	59	50	44	.014
	2. Answered correctly on discrimination index	24	41	50	36	.312
	3. Answered correctly on distractors and discrimination index	5	13	13	10	.628
Behavior ($p < .001$)	1. Conducts MCQ item analysis	33	72	63	57	.020
	2. Conducts MCQ item analysis more than 70% of departmental exams	19	34	38	30	.424
	3. Conducts MCQ item analysis more than 70% of module exams	14	59	50	43	.005
	4. Uses MCQ item analysis reports to modify exams	48	78	75	67	.061
FOM policy ($p = .838$)	1. Believe that there are strict departmental policies for applying MCQ item analysis	52	34	13	38	.120
	2. Believe that there are strict module policies for applying MCQ item analysis	33	63	75	54	.051

[a] *p* Values taken from pooled responses and based on χ^2 test.
[b] Values represent column percent.
[c] *p* Values taken from χ^2 test.

College policy on item analysis

Results reflect stricter policies for MCQ item analysis in module exams than in departmental exams, in which 54% of total participants believed that there was a strict policy within their module for applying MCQ item analysis, while 38% believed the same for departmental exams. There were no significant differences between the number of FDPs participants and the FOM policies, both at the departmental or at the module levels.

Discussion

The role of FDPs has gradually increased to accommodate multiple roles expected of faculty members. A continuous evaluation of such FDPs is needed to ensure that its ultimate goals can be achieved, and more importantly, sustained over longer periods of time. The objectives of the FDP MCQ item analysis were to support faculty to identify, interpret, value the contribution of item analysis, as it directly impacts the validity of the assessment. In this study, our results show the long-term effectiveness of MCQ item analysis training sessions provided by the KAU-FOM in improving the knowledge and the performance of faculty members. Findings indicate that attending FDPs on MCQ item analysis within the past 3 years increased confidence, knowledge, and behavior, including the application of item analysis to modify and improve test quality. The study included participation of different gender, academic positions, and years of teaching experience, based on a representative sample of participants. While prior studies have focused on short-term impacts or pre- and post-test changes, these longer-term studies, using a representative sample of faculty members provide a contribution to the literature.

Our findings are in agreement with other reports that have evaluated the utility of FDPs to improve the quality of MCQ item writing. In a study evaluating two training programs, researchers showed significant improvements in the difficulty index and students' mean scores through post-workshop training (Abdulghani et al. 2015). In this study, the authors measured the quality of pre-training and post-training MCQs construction of three courses. A major difference between our study and work of Abdulghani and colleagues is that the latter assessed the fourth level of the Kirkpatrick's model, which evaluated the change among the participants' performance in MCQs items' writing in exams. However, their assessment was conducted within the same year of the workshop training, while our study assessed the impact of FDPs within the last 3 years. Similarly, Jozefowicz et al. (2002) found that formal training in MCQ writing improved the quality of exam questions, when analyzing the United States Medical Licensing Examination (USMLE) Steps 1 and 2 exams (Jozefowicz et al. 2002).

In our study, it was clear that simply attending more workshops on MCQ item analysis did not necessary enhance the knowledge or the performance for specific traits. This was also reflected in the open ended narrative comments in the questionnaire by faculty members, in which some discussed their disapproval on attending repeated training sessions that would address the same topic, but could be misleading by the title of the training session. However, we found that FDPs, in their long-term impact, do have a significant systematic effect on the overall perception, knowledge, and behaviors of faculty members. Moreover, this finding was shown in a non-North American or European setting, indicating the potential benefit that FDPs can have in other Saudi medical schools and in the region.

Interestingly, however, when assessing the change in practice, we noticed a difference in practicing MCQ item analysis between departmental exams and in module exams, with almost 1.5 times the number reporting to use MCQ item analysis more than 70% in module exams. This could reflect differences in implementation of policies, since almost double the number of participants believed that there were more strict policies for applying MCQ item analysis in module exams than departmental exams. Our results thus suggests that better implementation of policies might be a crucial factor for enforcing the beneficial effects of FDPs.

When assessing the knowledge of participants on MCQ item analysis parameters, it was clear that participants knew more about the difficulty or the discrimination indices, with an average of 40% answering correctly questions related to these parameters, but had less knowledge on assessing a distractor, with only 10% answering correctly. When addressing the thoughts of participants on MCQ item analysis report parameters, many reported to rely heavily on the difficulty index of an item, using it to eliminate a question or to keep it, while placing minimal attention to the quality of distractors. We believe that more attention and analysis of the MCQ item distractors can enhance the quality of exam items. Paying attention to the quality of the distractors can assess in modifying items rather than completely eliminating them, which would save more energy and time for future exams. Our findings also note that despite 92% of participants reporting prior knowledge of MCQ item analysis, there were significant differences in their actual reported knowledge in the components of item analysis—and as such, these findings underscore discrepancies in self-perceived knowledge of MCQ item analysis versus actual knowledge.

Many studies have evaluated the effectiveness of FDPs immediately after the program (Leslie et al. 2013; Abdulghani et al. 2014). Our study is the first study to address the long-term benefits of MCQ item analysis training programs provided by the FOM. There are some limitations to this study. When assessing the practice and performance changes, our study used self-reported data from participants; however, we believe that further analysis of empirical exams results, using pre- and post-training sessions, in addition to students' mean scores and a comprehensive set of psychometric analyzes (reliability and item analysis), could provide a more objective indicator on the change in practice achieved by the training session. Efforts are underway to combine these data to supplement these findings in a future study. In addition, while our study examined basic participant faculty characteristics, other factors such as faculty interest, qualifications, and specific time of FDP attendance (to inform recall or maturity bias) could have influenced the results; and as such, we plan to further our understanding by examining other features of faculty that may affect their confidence, knowledge, and behaviors in future research. Moreover, our study is based on a single

institution. However, the mix of faculty junior and senior faculty members and the teaching experience may help to increase the value of our findings. These results also highlight specific targeted areas of workshop agenda and instruction that may continue to improve the effectiveness of the FDP. For example, specific areas in participant confidence, such as belief in benefits of the training and the impact of quality assurance on student learning, can be targeted for further emphasis in future FDPs, as these features may support added value to the workshops.

In conclusion, findings from this study suggest that FDPs, particularly focused on MCQ item analysis, could yield long-term systematic change on the confidence, knowledge, and behaviors of faculty members. As such, promoting FDPs can help to support faculty. Moreover, FDPs also need support from the departmental committee and a follow-up strategy to ensure its full effectiveness.

Acknowledgements

The authors thank members of the Medical Education Department at KAU, Nasra Ayoub, and Reda Jamjoom for supplying the relevant information related to MCQ item analysis training sessions provided by the department.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article..

Funding

The publication of this supplement has been made possible with the generous financial support of the Dr. Hamza Alkholi Chair for Developing Medical Education in KSA.

Notes on contributors

Aliaa Amr Alamoudi, MD, PhD, is an Assistant Professor in the Department of Clinical Biochemistry at King Abulaziz University (KAU). Dr. Alamoudi conceptualized and designed the study, acquired and interpreted the data, drafted the initial manuscript, revised the manuscript, and approved the final manuscript.

Basem Salama El-Deek, MD, PhD, is Professor in Community Medicine and in Medical Education departments in the Faculty of Medicine at KAU. Dr. El-Deek contributed to the conceptualization and revision of the manuscript.

Yoon Soo Park, PhD, is Assistant Professor in the Department of Medical Education at the University of Illinois College of Medicine at Chicago. Dr. Park contributed to the conceptualization and revision of the manuscript.

Lana Adey Al Shawwa, PhD, is Associate Professor in the Department of Medical Education in the Faculty of Medicine at KAU. Dr. Al Shawwa contributed to the conceptualization and revision of the manuscript.

Ara Tekian, PhD, MHPE, is Professor in the Department of Medical Education at the University of Illinois College of Medicine at Chicago. Dr. Tekian contributed to the conceptualization and revision of the manuscript.

ORCID

Yoon Soo Park http://orcid.org/0000-0001-8583-4335

References

Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK, Syed S, Aldrees AA, Alrowais N, Haque S. 2015. Faculty development programs improve the quality of Multiple Choice Questions items' writing. Sci Rep. 5:9556.

Abdulghani HM, Shaik SA, Khamis N, Al-Drees AA, Irshad M, Khalil MS, Alhaqwi AI, Isnani A. 2014. Research methodology workshops evaluation using the Kirkpatrick's model: translating theory into practice. Med Teach. 36:S24–S29.

Ayuob N, Eldeek B, ElHak GS, Jamjoom R. 2015. Enhancing medical curriculum integration from theory to practice; experience of two Arabic medical schools. Sylwan. 158:257–261.

Hochlehnert A, Brass K, Moltner A, Schultz JH, Norcini J, Tekian A, Junger J. 2012. Good exams made easy: the item management system for multiple examination formats. BMC Med Educ. 12:63.

Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. 2002. The quality of in-house medical school examinations. Acad Med. 77:156–161.

Kirkpatrick D. 1967. Evaluation of training. In: Craig R, Bittel L, editors. Training and development handbook. New York (NY): McGraw-Hill; p. 131–167.

Leslie K, Baker L, Egan-Lee E, Esdaile M, Reeves S. 2013. Advancing faculty development in medical education: a systematic review. Acad Med. 88:1038–1045.

McAndrew M, Motwaly S, Kamens TE. 2013. Long-term follow-up of a dental faculty development program. J Dent Educ. 77:716–722.

Moss E. 2001. Multiple choice questions: their value as an assessment tool. Curr Opin Anaesthesiol. 14:661–666.

Moussa MA, Ouda BA, Nemeth A. 1991. Analysis of multiple-choice items. Comput Methods Programs Biomed. 34:283–289.

Naeem N, van der Vleuten C, Alfaris EA. 2012. Faculty development on item writing substantially improves item quality. Adv Health Sci Educ Theory Pract. 17:369–376.

Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. 2006. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. Med Teach. 28:497–526.