

# A One-Day Dental Faculty Workshop in Writing Multiple-Choice Questions: An Impact Evaluation

Eiad AlFaris, MMed; Naghma Naeem, PhD; Farhana Irfan, MRCP;  
Riaz Qureshi, FRCGP; Hussain Saad, MRCP; Ra'ed Al Sadhan, MS;  
Hamza Mohammad Abdulghani, FRCGP; Cees Van der Vleuten, PhD

**Abstract:** Long training workshops on the writing of exam questions have been shown to be effective; however, the effectiveness of short workshops needs to be demonstrated. The aim of this study was to evaluate the impact of a one-day, seven-hour faculty development workshop at the College of Dentistry, King Saud University, Saudi Arabia, on the quality of multiple-choice questions (MCQs). Kirkpatrick's four-level evaluation model was used. Participants' satisfaction (Kirkpatrick's Level 1) was evaluated with a post-workshop questionnaire. A quasi-experimental, randomized separate sample, pretest-posttest design was used to assess the learning effect (Kirkpatrick's Level 2). To evaluate transfer of learning to practice (Kirkpatrick's Level 3), MCQs created by ten faculty members as a result of the training were assessed. To assess Kirkpatrick's Level 4 regarding institutional change, interviews with three key leaders of the school were conducted, coded, and analyzed. A total of 72 course directors were invited to and attended some part of the workshop; all 52 who attended the entire workshop completed the satisfaction form; and 22 of the 36 participants in the experimental group completed the posttest. The results showed that all 52 participants were highly satisfied with the workshop, and significant positive changes were found in the faculty members' knowledge and the quality of their MCQs with effect sizes of 0.7 and 0.28, respectively. At the institutional level, the interviews demonstrated positive structural changes in the school's assessment system. Overall, this one-day item-writing faculty workshop resulted in positive changes at all four of Kirkpatrick's levels; these effects suggest that even a short training session can improve a dental school's assessment of its students.

Prof. AlFaris is Professor of Family Medicine and supervisor of King Saud University Chair for Medical Education Research and Development, Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia; Prof. Naeem is Associate Professor and Head of Medical Education Department, Batterjee Medical College, Jeddah, Saudi Arabia; Prof. Irfan is Assistant Professor and Chair of Medical Education Research and Development, Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia; Prof. Qureshi is Distinguished Professor of Family Medicine, Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia; Prof. Saad is Assistant Professor, Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia; Prof. Al Sadhan is Associate Professor and Head of Oral Medicine and Diagnostic Sciences, College of Dentistry, King Saud University, Riyadh, Saudi Arabia; Prof. Abdulghani is Associate Professor and Head of Assessment and Evaluation Centre, College of Medicine, King Saud University, Riyadh, Saudi Arabia; and Prof. Van der Vleuten is Professor of Education, Scientific Director of School of Health Professions Education, Faculty of Health, Medicine, and Life Sciences, Department of Educational Development and Research, Maastricht University, Netherlands. Direct correspondence to Prof. Farhana Irfan, Department of Family and Community Medicine (34), College of Medicine, King Saud University, P.O. Box 2925, Riyadh 11461, Kingdom of Saudi Arabia; +966114690328; masoodfarhana1969@gmail.com.

**Keywords:** dental education, assessment, multiple-choice questions, dental faculty, faculty development, Kirkpatrick's framework

*Submitted for publication 12/30/14; accepted 5/2/15*

Multiple-choice questions (MCQs) are commonly used in the assessment of students in the health professions. If constructed properly, MCQs can test a large number of learners, can cover extensive content areas, and are associated with better grading accuracy and consistency than subjective forms of assessment.<sup>1</sup> Furthermore, well-designed MCQs can test students' knowledge, its application, and problem-solving cognitive skills.<sup>2</sup> Unfortunately, most instructors fail to follow generally accepted principles of MCQ construction<sup>3-5</sup> and test mainly the lower cognitive levels as defined in

Bloom's taxonomy.<sup>6-8</sup> Many examinations therefore suffer from lack of quality.<sup>3</sup>

Ensuring high quality of test items is a challenge, especially without faculty development. A systematic review by Steinert et al. in health professions education found faculty training for teaching effectiveness to be associated with positive changes in both the short term and the long term.<sup>9</sup> However, in relation to faculty development specifically related to item-writing, very few studies have been done. One study in a medical school that involved a long training workshop (one full week) found an increase

in faculty confidence and skill in writing high-quality MCQs and short-answer questions.<sup>10</sup> The extensive training program in that study raises feasibility questions in relation to practical implications.

In dental schools in the U.S., “ongoing faculty development is a requirement” in the Commission on Dental Accreditation (CODA) standards in order “to improve teaching and learning, to foster curricular change, to enhance retention and job satisfaction of faculty, and to maintain the vitality of academic dentistry as the wellspring of a learned profession” (p. 12).<sup>11</sup> Dalrymple et al.’s two studies and Hendricson et al.’s are among the small number of studies of the outcomes of faculty development programs in dental schools.<sup>12-14</sup> Another study found training to be associated with positive long-term changes in teaching behaviors.<sup>15</sup> A study by Meyari et al., which assessed the quality of MCQs in annual residency exams before and after a workshop, found some improvement in the quality of the MCQs written after the training.<sup>16</sup>

The aim of our study was to assess the impact of a one-day MCQ-writing workshop on the quality of in-house MCQ questions in a dental school. Kirkpatrick’s four-level evaluation framework was used to evaluate the participants’ reactions to the workshop (Level 1), their learning (Level 2), their behavioral change (measured by the quality of their MCQs after the workshop) (Level 3), and the workshop’s impact (measured by institutional change) (Level 4).<sup>17</sup>

## Methods

The study was reviewed by the Institutional Review Board of King Saud University and was found to be ethically acceptable. The study was conducted from the end of 2012 to the beginning of 2013 at the College of Dentistry, King Saud University.

The seven-hour, one-day workshop was held for two cohorts of course directors in the undergraduate program. At the beginning of the workshop, the learning issues were described, and the expectations of the participants were ascertained and written on a flip chart. Sessions lasted from 15 to 60 minutes (Table 1).

Six teaching methods were used to achieve the workshop objectives. The first were short interactive lectures that focused on the important concepts. The presenters made sure that the participants understood the basic themes. Second was demonstration of common flaws in MCQs. Third were small-group exercises in which the participants were encouraged

to perform specific tasks, such as identification of pitfalls in writing MCQ items. They were also provided with templates as a stimulus and guide for the group discussion. Fourth was feedback. The participants were asked to make presentations of their products in small-group exercises. The presenting groups made positive points about their own performance, and members of the other groups were asked to add additional positive points. The same procedure was followed for points regarding the need for improvement, after which one of the facilitators provided a summary. Fifth was discussion on interpretation of item analysis and psychometric reports followed by, sixth, discussion of how to make a positive change at the institutional level. At the end of the workshop, a 15-minute discussion was held on how to ensure improvement of the quality of MCQs and to sustain that improvement.

## Quantitative Assessment

Immediately after the workshop, the participants who had attended the entire training day were invited to fill out a satisfaction questionnaire to evaluate Kirkpatrick’s Level 1 (reaction). The questionnaire solicited feedback on content, instructional design, facilitation, and outcome. Responses were selected on a scale of 1-5 with 1=lowest and 5=highest. No open-ended questions were used.

To assess Level 2 (learning), a quasi-experimental design with randomized separate samples pretest and posttest was used.<sup>18,19</sup> The tests consisted of 20 MCQs and short answer questions regarding aspects of assessment and using MCQs. Before the workshop, the pretest was administered to a random half of the participants (control group), while the other half of the participants filled out a demographic and general habits questionnaire (experimental group). At the end of the workshop, the control group filled out the demographic and general habits questionnaire, and the experimental group filled out the posttest (Figure 1). The reason for using this technique was that it reduces the bias associated with learning in a pretest-posttest design since test takers in the experimental group are not exposed to the questions before the intervention.<sup>18</sup>

Ten of the participants from the experimental group were randomly selected for the Level 3 (behavioral changes) part of the study. Each of these faculty members provided 50-100 MCQs from the annual examination they had developed and administered before the workshop and a similar number they wrote

**Table 1. Schedule of one-day faculty workshop on writing multiple-choice questions (MCQs)**

Time	Activity	Type of Activity
7:30-8:00	Registration	
8:00-8:30	Introduction	Large group
	Your expectations of this workshop	Participants' input
	Workshop schedule: how it fits with your expectations	Facilitator feedback
8:30-9:00	Assessment purposes and types	Short interactive lecture
9:00-9:30	Common pitfalls in MCQ writing	Short interactive lecture
9:30-10:00	Exercise on identifying pitfalls in real MCQs	Small-group exercise
10:00-10:30	Steps in writing MCQs	Facilitator presentation
10:30-11:00	Develop MCQs	Small-group activity
11:00-11:30	Presentation of MCQs presented by small-group rapporteurs	Review and critique in large group
11:30-12:00	Develop high-quality MCQs using quality checklists	Small-group exercise
12:00-1:00	Lunch and prayer	
1:00-1:15	Introduction to blueprinting	Short presentation
1:15-2:00	Develop your test blueprint	Small-group exercise and feedback
2:00-3:00	Review of what was done	Large-group discussion
3:00-3:30	Psychometric analysis	Short presentation
3:30-4:00	Psychometric analysis	Small-group exercise and feedback

six months after the workshop. Ten MCQs per faculty member were randomly selected from both the pre- and post-intervention exams. Therefore, a total sample of 100 MCQs from pre-training and another 100 from post-training were used for quality analysis.

Two of the authors (NN and EA), who are well versed in MCQ assessment, rated the quality of each of the 200 faculty-generated MCQs independently. The definition of item pitfalls was adopted from Case and Swanson's guidelines,<sup>5</sup> and the rating criteria were similar to those used by Jozefowicz et al.<sup>3</sup> Discrepancies were resolved through consensus. The two raters were not blinded to item origin (pre- or posttest). On a five-point rating scale, the items were rated as follows: 5=item included a vignette (either patient or laboratory), a one-best-answer format (i.e., not true or false), and no technical flaws that provided special benefit to testwise examinees or that posed irrelevant difficulty;<sup>4,5</sup> 4=item failed to completely satisfy one of the three conditions in a five-point item; 3=item did not involve a vignette but was focused, contained no item flaws, and involved more than recall of an isolated fact; 2=item failed to satisfy one of the criteria for a three-point item; and 1=item failed to satisfy two or all three of the criteria for a three-point item.

## Qualitative Assessment

Kirkpatrick's Level 4 (organizational impact) was evaluated by interviewing three key administrators of the school. Furthermore, the ten faculty

members who participated in the Level 3 evaluation were asked by email what they did in the six months between the pre- and posttests to develop their MCQ writing skills.

The interview protocol was formulated by brainstorming the main research issues, led by two qualitative research experts (NZ and NZ) along with coauthors EA and FI. The protocol was designed to capture four Ws (what, why, when, and who) in order to elicit the big picture of the overall process of change at the organizational level. The interviews were initiated by providing basic information about the study to the college's former dean, current academic vice dean, and former vice dean and soliciting their consent to be interviewed. Data collection was done through semistructured interviews, which allowed an in-depth exploration of the research issue guided by the interview protocol. The length of the interviews ranged from 40 to 50 minutes. Interviews were recorded using digital recorders. The principal investigator (EA) conducted all interviews in the presence of one coauthor (FI). The interviews were transcribed verbatim by FI manually.

The informants were asked about institutional changes that occurred in 2011-14 in chronological order to make it easier to follow the flow of their stories. Cross-checking with the interviewee was conducted immediately after the interview to verify accuracy of the transcribed data. All the transcriptions were uploaded on Atlas.ti 7.2. The first step was to read all the transcriptions and create open coding.

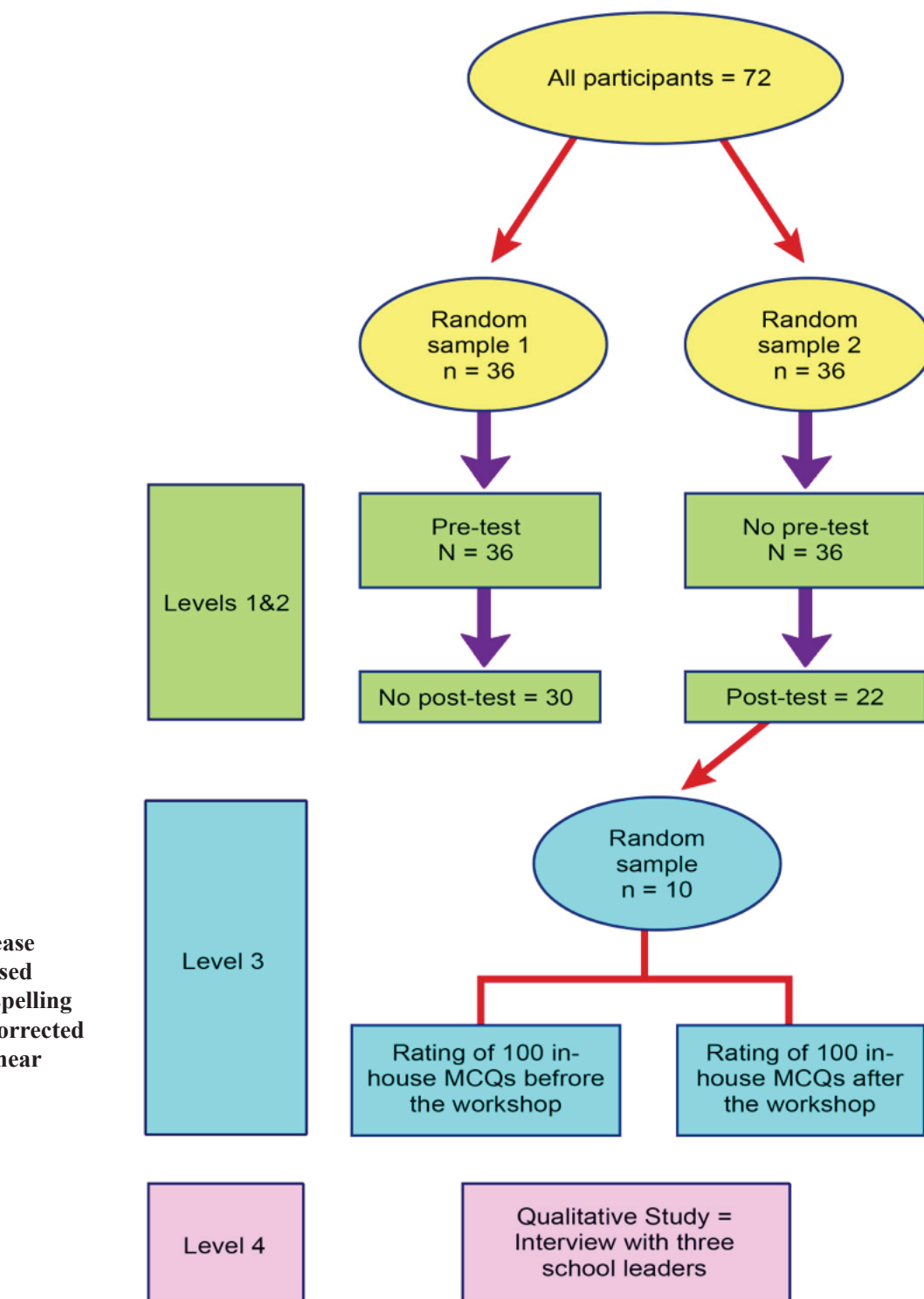


Figure 1. Overview of study design, organized by Kirkpatrick's Levels 1, 2, 3, and 4

Codes were created using a word/term from the protocol such as “reasons for change” and “benefits of change.” We also recoded some responses to specific subcodes, such as “student feedback” and “student complaints,” to represent subtopics under “reasons for change” and create other subcodes. The coding process was discussed with one of the two coauthor experts to ensure accuracy.

## Statistical Analyses

Descriptive statistics were calculated for Kirkpatrick’s Level 1. For Level 2, changes in score and effect size were calculated using Cohen’s *d*.<sup>20</sup> For Level 3, the mean of the two assessors’ ratings of the MCQ quality score for each item, along with the change in score analysis and effect sizes, was calculated to compare the pretest and posttest results. For Level 4, the interview transcript was summarized and interpreted.

## Results

An almost equal number of male and female faculty members participated in some part of the workshop (N=72) (Table 2). However, only participants who attended the complete activity (n=52) were included in Level 2 of the study; one-third of them were women, while two-thirds were men.

### Level 1: Reaction

Of the total 72 course directors, 52 attended the entire workshop (30 from sample 1 and 22 from sample 2) (Figure 1). All 52 completed the satisfaction questionnaire for a response rate of 100%. Overall, the participants responded favorably to the training. Their average ratings on the satisfaction scale for teaching content, instructional design, facilitation, and achievement of goals were high (4-5) on a scale of 1-5 (Figure 2). The workshop facilitators felt that the small-group exercises complemented with feedback were the most useful part of the workshop.

### Level 2: Learning

In this phase of the study, 30 participants in the control group (n=36) completed the pretest (response rate 83%); and 22 participants in the experimental group (n=36) completed the posttest (response rate 61%). The results showed the control group’s knowledge test score before the workshop was low, with

**Table 2. Characteristics of total study population (n=72)**

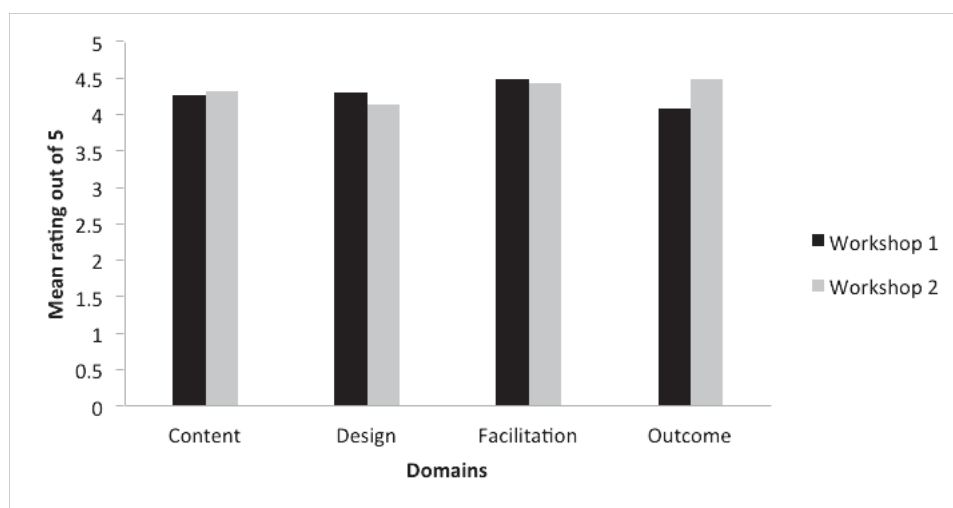
Characteristic	Number
Gender	
Male	35
Female	37
Specialty	
Restorative dental sciences (operative dentistry, endodontics, dental biomaterial sciences)	23
Diagnostic dental sciences (oral biology, oral microbiology, oral diagnosis, oral medicine, oral pathology, oral radiology)	15
Prosthodontics	14
Pediatric dentistry and orthodontics	6
Periodontics and community dentistry	5
Oral and maxillofacial surgery	8
Dental hygiene	1
Educational rank	
Lecturer	16
Senior registrar	1
Assistant professor	23
Associate professor	19
Professor	13
Years of teaching	
Less than 5	26
5-10	11
11-15	11
>15	24

a mean ( $\pm$ SD) of  $4.56 \pm 2.72$  out of a maximum of 20. The experimental group’s mean score after the workshop was  $12 \pm 4.7$ . This reflected a large effect size of 0.72 (Table 3).

### Level 3: Behavioral Change

The percentage of agreement between the two raters of the participants’ MCQs was 90%. The quality of the MCQs before training was low with a mean ( $\pm$ SD) of  $2.27 \pm 1.14$  out of a possible score of 5; the corresponding figure for post-training was higher, at  $2.87 \pm 0.79$ . Mean scores and effect size demonstrated improvement in the quality of MCQs with a small to moderate effect (effect size 0.28) (Table 4). When the ten faculty members who participated in the Level 3 part of the study were asked what they did in the six months between tests, eight said they discussed their MCQs with colleagues/seniors or asked them to review their MCQs. Three mentioned that they had formal training, and three stated that they did additional reading.





**Figure 2.** Participants' evaluation of four aspects of faculty development workshop by mean rating on scale from 0=low to 5=high

**Table 3.** Participants' pretest-posttest results regarding knowledge (mean and standard deviation)

Number of Participants	Pretest Mean	Posttest Mean	Effect Size
22	4.56±2.72	12±4.7	0.72

*Note:* The maximum score for both the pretest and the posttest was 20. Effect size was large according to Cohen's classification;  $p < 0.05$ .

**Table 4.** Quality scores (mean and standard deviation) of participants' multiple-choice questions (MCQs) before training workshop and six months after

Number of Participants	Mean MCQ Rating Scores Before Training	Mean MCQ Rating Scores After Training	Effect Size
10	2.27±1.14	2.87±0.79	0.28

*Note:* Maximum possible score was 5. Effect size was small according to Cohen's classification;  $p > 0.05$ .

## Level 4: Impact

Four major themes were found to support the impact of the workshop at the institutional level. Our discussion of these themes presents them according to the frequency of mentions by the interviewees.

Reasons for change emerged as one of the top themes with the highest number of codes. Comments falling under this theme were mostly students' concern and complaints about MCQs used in exams, improvement of services for the students, problems in assessment, and the need to change as part of the accreditation process. One interviewee commented, "There were complaints of the students regarding the types of questions they were getting in exams.

When we reviewed some samples, we recognized that there was a problem in the construction of the multiple choice questions." According to another, "It started as part of preparation for the accreditation."

Regarding the planning and change process, administrative support was related to the dean's verbal instruction to the teaching staff during the workshop and when visiting all the departmental boards. The school leadership distributed a free copy of a booklet<sup>21</sup> to all the faculty members. Furthermore, the dean made it mandatory that all the departments must establish departmental review committees.

Challenges were derived as a theme from a series of comments by the three interviewees in re-

sponse to two questions: Did you face any challenges to bring about the change? If yes, what were they? Comments in response to these questions provided an overview of factors, such as non-acceptance of faculty members to send their exam questions for review by colleagues, late submission of questions for review by other course faculty members, and unwillingness to review the questions as it leads to extra workload. One interviewee noted that “some faculty, particularly course directors or seniors, think that there is no need to send the questions to the review committee as it is their course, their subject so they are the ones who know the course objectives and they know how to structure the exam.” This resistance was overcome by communication at the personal level, as well as using electronic means to address the issue and by explaining both existing and upcoming strategies for improvements at the organizational level to achieve the goal.

The interviewees were asked to contextualize their feelings, experiences, and beliefs about the impact of changes that took place. Statement after statement, the interviewees referred to the positive changes, such as fewer complaints by students about the exams and greater student satisfaction. Comments revolved around the positive feedback from students as well as the faculty regarding a better exam structure. One interviewee noted that “little complaint comes from students regarding difficulty of the questions,” and another stated, “Now we have better MCQ exams and the amount of rejection of questions by the review committee has reduced.” The third interviewee commented, “Students feel the change in the quality of questions given in the exams, and we have also received positive feedback from the students’ leaders during our meetings.”

---

## Discussion

In this study, the impact of a relatively short MCQ construction workshop on the quality of test items was evaluated using Kirkpatrick’s framework.<sup>17</sup> The participants’ satisfaction with the workshop elements was encouraging and is consistent with the findings of a systematic review of faculty development studies, which reported that workshops were generally rated as helpful, relevant, and useful and were very well received.<sup>9</sup> Although satisfaction with a training program is an elementary level of evaluation, it is fundamental for a change to occur.<sup>22</sup> The substantial improvement in the participants’ knowl-

edge score (learning) after the training intervention is consistent with previous studies that found positive changes in teachers’ knowledge, attitudes, and skills in teaching effectiveness following participation in a faculty development activity.<sup>9,19</sup>

Behavioral change (Level 3) is not frequently investigated<sup>23</sup> although it is an essential component of any training exercise.<sup>24</sup> This level investigates maintenance of the skills learned in the workshop over a period of time, leading to transfer of learning to practice. Therefore, the fact that a small to moderate effect size was found in the quality of MCQs six months after the training course is encouraging. It opens the space for further research to explore efficient and perhaps more cost-effective instructional methods when conducting MCQ writing workshops, e.g., trying blended formats in online and face-to-face components or a series of workshops. One study suggested that item writing could be mastered through extensive and critically supervised practice and protected time.<sup>25</sup> However, it is difficult for many faculty members to attend lengthy training sessions.<sup>26</sup> Previous research shows that, contrary to short courses, long courses may have a more significant impact on behavior regarding quality of MCQ exams.<sup>10,27,28</sup>

Our study found that a short workshop provided demonstrable effectiveness and had a positive impact on transfer of knowledge and skills. Repeated interventions, known as “booster sessions,” have been shown to enhance cumulative learning, practice, and maintenance of positive change over time.<sup>28,29</sup> Features mentioned in one study to be associated with effectiveness were the use of experiential learning, provision of feedback, effective peer and colleague relationships, well-designed interventions following principles of teaching and learning, and the use of multiple methods of teaching and learning within single interventions.<sup>9</sup>

While little previous research has evaluated Kirkpatrick’s Level 4 by examining the change in organizational practice,<sup>30-32</sup> our study attempted to explore this area. The establishment of an institutional review process to ensure the quality of all MCQs before being included in final examinations is a positive change that may complement the behavioral change. It is rather remarkable that sustained positive institutional change took place as a result of a single-day workshop. One explanation for the positive changes is the institutional buy-in, as evidenced by the commitment of the leadership. Furthermore, having a discussion with the participants during the workshop and later with the school’s leaders on their

assessment strategy and its importance for the quality of the training program had a positive impact. As the feedback was positively received by the trainees, we strongly encourage including such feedback in future training programs. A systematic review by Davis et al. found that feedback is helpful for transfer of learning to take place.<sup>28</sup> Furthermore, we suggest that, for MCQ writing improvement to be sustainable, it is important to include individual feedback by educators. This recommendation is consistent with Kotter and Cohen's statement that the core of the matter for helping people to change their behavior is never the strategy, structure, culture, or systems, but rather by addressing people's feelings.<sup>33</sup>

Future research should focus on constructs, such as individual, motivational, and environmental factors that affect transfer of knowledge and skills to elaborate on why some faculty members sustained the learning from workshop and were able to transfer it to the workplace, while others did not. Models exploring the relationship and interplay between these factors can advance our knowledge about transfer of learning.<sup>34</sup> This insight, in turn, can be used to design training programs that optimize and sustain transfer of learning from workshops to workplace.

This study had a number of limitations. The participants were from one dental school and limited in numbers and length of follow-up. The two raters were not blinded to MCQ origin (pre- or posttest). The students' satisfaction was not measured formally, which could have added to the accuracy of Kirkpatrick's Level 4 evaluation. We cannot entirely attribute the changes in the MCQs' quality to attendance at the workshop as there was no overall control group consisting of those who did not attend the workshop for the Level 3 evaluation. Another limitation is that only 22 participants (out of 36 assigned to take the posttest) completed the posttest. The strengths of the study are the evaluation across Kirkpatrick's four levels, the use of a quasi-experimental design with randomized separate samples, pre- and posttests for Level 2, and the triangulated design involving quantitative (Levels 2 and 3) and qualitative (Level 4) components.

## Conclusion

The results of this study suggest that a short (single-day) workshop may lead to a large effect size on learning and knowledge gain and a small to moderate effect on the quality of MCQs written by

faculty members. A significant sustained change in the institutional assessment strategy was also observed. Additionally, institutional buy-in and leadership support after the main intervention were found to be important factors for sustaining the change.

## Acknowledgments

The project was supported by King Saud University, Vice Deanship of Research Chairs. We are thankful to pharmacist Abdullah M.A. Ahmed for his help in data collection. We are thankful to Dr. Nasriah Zakaria and Dr. Norhayati Zakaria for the qualitative part of the study and Dr. Gominda Ponnampuruma for reviewing the manuscript before submission.

## REFERENCES

1. Case S. Strategies for performing well on multiple-choice question tests. In: Al Alwan I, Magzoub ME, Elzubeir M, eds. *An international handbook of medical education: a guide for students*. London: Sage, 2012:247-53.
2. Abdel-Hameed AA, Al-Faris EA, Alorainy IA. The criteria and analysis of good multiple-choice questions in a health professional setting. *Saudi Med J* 2005;26(10):1505-10.
3. Jozefowicz RF, Koeppen BM, Case S, et al. The quality of in-house medical school examinations. *Acad Med* 2002;77(2):156-61.
4. 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(2):133-43.
5. Case S, Swanson D. *Constructing written test questions for the basic and clinical sciences*. 3rd ed. Philadelphia: National Board of Examiners, 2003.
6. Gruppen L, White C. Identifying learners' need and self-assessment. In: Dornan T, Mann K, Scherpbier A, Spencer J, eds. *Medical education theory and practice*. London: Churchill Livingstone, Elsevier, 2010:115-27.
7. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Med Educ* 2008;42(2):198-206.
8. Baig M, Ali SK, Ali S, Huda N. Evaluation of multiple-choice and short essay question items in basic medical sciences. *Pak J Med Sci* 2014;30(1):3-6.
9. Steinert Y, Mann K, Centeno A, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Med Teach* 2006;28(6):497-526.
10. Naeem N, Van der Vleuten CPM, Alfari EA. Faculty development on item writing substantially improves item quality. *Adv Health Sci Educ* 2012;17(3):369-76.
11. Commission on Dental Accreditation. Accreditation standards for dental education programs. At: [www.ada.org/~media/CODA/Files/predoc\\_2013.ashx](http://www.ada.org/~media/CODA/Files/predoc_2013.ashx). Accessed 30 Dec. 2014.
12. Dalrymple KR, Wuenschell C, Shuler CF. Development and implementation of a comprehensive faculty



- development program in PBL core skills. *J Dent Educ* 2006;70(9):948-55.
13. Dalrymple KR, Wuenschell C, Rosenblum A, et al. PBL core skills faculty development workshop 1: an experiential exercise with the PBL process. *J Dent Educ* 2007;71(2):249-59.
  14. Hendricson WD, Anderson E, Andrieu SC, et al. Does faculty development enhance teaching effectiveness? *J Dent Educ* 2007;71(12):1513-33.
  15. McAndrew, Motwaly S, Kamens TE. Long-term follow-up of a dental faculty development program. *J Dent Educ* 2013;77(6):716-22.
  16. Meyari A, Biglarkhani M, Zandi M, et al. The effect of education on improvement of design of multiple-choice questions in annual residency exams of dental school. *Iran J Med Educ* 2012;12(1):36-45. [in Persian]
  17. Kirkpatrick DL. Evaluating training programs: the four levels. San Francisco: Berrett-Koehler, 1998.
  18. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand McNally College Pub. Co., 1966.
  19. Markert RJ, O'Neill SC, Bhatia SC. Using a quasi-experimental research design to assess knowledge in continuing medical education programs. *J Contin Educ Health Prof* 2003;23(3):157-61.
  20. Hojat M, Xu G. A visitor's guide to effect sizes: statistical significance versus practical (clinical) importance of research findings. *Adv Health Sci Educ* 2004;9(3):241-9.
  21. AlFaris EA. MCQs for health professionals assessment: construction, evaluation, and interpretation. Riyadh: King Saud University, 2011.
  22. Belfield C, Thomas H, Bullock A, et al. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach* 2001;23(2):164-70.
  23. Abdulghani HM, Shaik SA, Khamis N, et al. Research methodology workshops evaluation using the Kirkpatrick's model: translating theory into practice. *Med Teach* 2014;36(1):S24-9.
  24. Baldwin T, Ford JK. Transfer of training: a review and directions for future research. *Personnel Psychol* 1988;41:63-105.
  25. Rodriguez MC. The art and science of item writing: a meta-analysis of multiple-choice item format effects. Paper presented at American Educational Research Association Annual Meeting, Chicago, IL, April 1997.
  26. Iramaneerat C. The impact of item writer training on item statistics of multiple-choice items for medical student examination. *Siriraj Med J* 2012;64(6):178-82.
  27. Mann KV. Thinking about learning: implications for principle-based professional education. *J Contin Educ Health Prof* 2002;22(2):69-76.
  28. Davis D, Evans M, Jadad A, et al. The case for knowledge translation: shortening the journey from evidence to effect. *BMJ* 2003;327(7405):33-5.
  29. Bland CJ, Froberg DG. A systematic approach to faculty development for family practice faculty. *J Fam Pract* 1982;14(3):537-43.
  30. Nathan RG, Smith MF. Students' evaluations of faculty members' teaching before and after a teacher-training workshop. *Acad Med* 1992;67(2):134-5.
  31. Sachdeva AK, Kelliher GJ. Impact of a faculty development workshop on three institutions. In: Rothman AI, Cohen R, eds. *Proceedings of the Sixth Ottawa Conference on Medical Education*. Ottawa: University of Toronto, 1994:75-9.
  32. Snyder S. A program to teach curriculum development to junior faculty. *Fam Med* 2001;33(5):382-7.
  33. Kotter JP, Cohen DS. The heart of change: real-life stories of how people change their organizations. *J Bus Behav Sci* 2002;26(2).
  34. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.