

Effect of the Educational Pamphlet on the Quality of Multiple-Choice Questions

F. Sayar¹, M. Vatanpour², M. Ghasemi¹, MR. Shabahangfar³✉.

¹ Associate Professor, Department of Periodontics, School of Dentistry, Dental Branch, Islamic Azad University, Tehran, Iran

² Assistant Professor, Department of Endodontics, School of Dentistry, Dental Branch, Islamic Azad University, Tehran, Iran

³ Assistant Professor, Department of Periodontics, School of Dentistry, Dental Branch, Islamic Azad University, Tehran, Iran

Abstract

Background and Aim: The aim of this study was to evaluate the structural integrity of multiple-choice questions according to the Millman checklist, and to assess the distinguishing power of these questions between weak and strong students in theoretical courses of Periodontics in an academic year (2014-15) in Dental Branch of Tehran Islamic Azad University.

Materials and Methods: A total of 262 students comprising periodontics module 1, 2 and 3 in two semesters responded to 240 multiple-choice questions and results from the two consecutive semesters were compared with each other. ANOVA was used to analyse data. $P < 0.05$ was considered significant.

Results: There was no statistical difference between the difficulty coefficient of the questions in the two semesters for Periodontics module 1, 2, and 3 (P -values were 0.2, 0.34 and 0.69, respectively). For discrimination coefficient, there was no significant difference between Periodontics module 2 ($P = 0.72$) and 3 ($P = 0.42$) in the two semesters, however, in Periodontics module 1 this difference was significant ($P = 0.017$). Although a favourable change in the indicators of Millman checklist was seen in the second semester, it was not statistically significant ($P = 0.65$). In total, the difficulty and discrimination coefficients revealed that the majority of questions were "medium difficult".

Conclusion: Although the educational pamphlet did not significantly improve the quality of multiple-choice questions, some Millman checklist items improved slightly.

Key Words: Millman checklist, Discrimination coefficient, Difficulty coefficient, Periodontics, Multiple choice questions (MCQ)

✉ Corresponding author:
MR. Shabahangfar, Assistant
Professor, Department of
Periodontics, School of
Dentistry, Dental Branch, Islamic
Azad University, Tehran, Iran

drshabahangfar@gmail.com

Received: 10 May 2016

Accepted: 25 Sep 2016

➤ Cite this article as: Sayar F, Vatanpour M, Ghasemi M, Shabahangfar MR. Effect of the Educational Pamphlet on the Quality of Multiple-Choice Questions. J Islam Dent Assoc Iran. 2016; 28(4):118-125. DOI: 10.30699/jidai.29.4.118

Introduction

One of the real concerns of the academic community, students and especially faculty members, is the students' assessment regarding the learning of theoretical content of academic courses [1,2]. For the first time in 1914, Fredrick Kelly raised the issue of quality of multiple-choice questions [3]. This kind of questions was studied five decades ago in the United States specialized boards written exams, for the first time [4]. The evaluation of compliance with the structural

principles in the multiple-choice questions, which are the correlation between the question stem and options, was studied by Millman in 1981 [3]. Study of the difficulty level and discrimination power was carried out by test developing professionals, Mehrens and Lehmann, at 1984 [5].

Miller proposed a "framework for clinical assessment", which recognizes that it is reasonable to evaluate different learning outcomes with different assessment methods in medical education. At the base of the pyramid is some guarantee that a

student “*knows*” what is required and essential in order to achieve efficacy in professional functions. The knowledge base is necessary and ought to be measured; however, knowledge alone is not enough to determine medical practice. Students should also show capability and competence by “*knowing how*” to use their accumulated knowledge and by “*showing how*” they do it during their performance. Finally, in order to demonstrate the “*action*”, students require to independently exhibit their professional behavior by “*doing*” or functioning in the real setting of clinical practice [6].

The learning results of action and performance are progressed productively based on other learning outcomes that should occur and develop first. In other words, proficiency in the lower levels of pyramid results in the quality of the performance outcomes, which is the higher phase in the hierarchy. Nevertheless, should bear in mind that the evaluation of knowledge and aptitude, which are lower in the triangle, cannot be presumed to fully predict the achievement of the more complex goals in the upper levels. In order to meticulously and realistically assess the different kinds of learning outcome, the use of various assessment tools is required [6].

The assessment of the capability of undergraduate dental and medical students is a very crucial task since these ‘to be physicians’ have to serve and save the human lives. Undergraduate students should be evaluated in three domains of educational activities or learning, i.e. Cognitive, Affective, and Psychomotor. Cognitive domain includes various levels of Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. Since the most physicians spend their time analyzing patient’s problems, the major emphasis in the medical education is on developing and evaluating the level III or problem-solving skills [1].

In an appropriate and precise cognitive assessment, students should be rewarded for their developed cognitive skills and intellectual thinking. The knowledge domain can be assessed by various methods which include Free response examinations (Long Essay Questions, Short Answer Questions, Modified Essay Questions), Multiple choice questions (MCQ), Key feature

questions, Self-assessment, and peer-assessment. All of these methods have their advantages and disadvantages. No single method of evaluation is better than the other and probably a reliable and valid evaluation requires a combination of two or more of these methods [1].

Teachers try to incorporate the best questions, more numbers of questions, and more options for better evaluation of students [7,8].

Multiple choice questions (MCQs) are one of the most frequently used methods for evaluating the theoretical knowledge of students and multiple studies have been conducted on their quality indicators, taxonomy of the questions, and the role of educational interventions on optimization of the multiple-choice questions [1,8,9]. One of the best methods to evaluate the MCQs is the use of Millman method [3].

The Multiple-choice questions are very common in the evaluation of undergraduate medical and dental students. They are reliable, valid, and easy to manage in a large number of students. Although well-designed MCQs have an extensive ability to test knowledge and factual recall, they are less effective in determining the problem-solving skills of students. Computer software can be used for easy and reliable scoring, however, design and creation of good MCQs is challenging and needs expertise [1]. If questions are drafted in a way that fail to measure students’ knowledge, studious and active students would be disillusioned and also it would cause more negligence in weak students and in consequence, the public health will be at risk.

Given the importance of in-service training for the faculty members of universities [8] and the presence of information gap about the quality of the exam questions, the purpose of present study was to evaluate multiple-choice questions of periodontics in the final exam, in terms of difficulty coefficient, discrimination coefficient and Millman standards, as well as the effect of education of the faculty members on this process at Department of Periodontology of Dental Branch of Tehran Azad University.

Materials and Methods

This action research was conducted as an analytical and experimental study to evaluate MCQs of periodontics modules 1, 2 and 3 exams

during two semesters in 2014-15 academic years in the theoretical unit. All MCQs had a single stem with four options/responses including, one correct answer and three incorrect alternatives. At the beginning of the second semester, faculty members were given a pamphlet explaining the basic principles in designing questions according to Millman checklist. All the questions for the first and the second semesters were designed by the same faculty members. The quality and quantity of the questions tried to be the same for the first and second semester exams.

Multiple-choice questions were evaluated according to the Millman checklist. Fourteen indicators were used including: stem clearness, specific objective of the question, negative option for the stem, specific option, contrasting option, positive words in the stem and options, structure of writing of the stem, duplicate option, spelling of stem and option, vertically writing of the options, positivity of the stem and options, use of "all items" and "none of the above" phrases in the options. Therefore, the final score of each question and consequently the total score of the questionnaire of that semester was determined.

Students answer sheets were evaluated and the scores of 25% of students with the highest rank and 25% of students with the lowest rank in the exams were collected [3]. Difficulty index and discrimination index were determined for each question. Difficulty coefficient was calculated as the percentage of the total number of people who correctly answered a question divided by the number of examinees and discrimination coefficient was calculated as the highest rank group right choices minus lowest rank group right choices divided by the number of people in a group (highest or lowest) [3].

At the end of the study, difficulty and discrimination coefficient of each question, correlation of total score with the answer of each question, and the percentage of responses for each question from the options available was determined. Statistical analysis was performed using SPSS 16 software and appropriate descriptive statistics methods, t-test analysis, and Pearson correlation coefficient were calculated.

Results

The present study was conducted on 262 students of Islamic Azad University, Dental Branch, in the academic year 2014-15 and 240 multiple choice questions in the final exam of theoretical course of Periodontology modules 1, 2 and 3 were analyzed. Since the sample size was more than 40, only 20% of the students with the highest score and 20% with the lowest marks were selected for statistical analysis [3]. Difficulty coefficients and discrimination coefficients are presented in Table 1 and 2, respectively. In module 1, the number of multiple-choice questions in the first and the second semester were 18 and 58, respectively. Discrimination coefficient of these questions were 0.38 ± 0.18 and 0.29 ± 0.18 for the first and the second semester respectively, which t-test revealed that the difference was statistically significant ($P=0.017$). The difficulty coefficient of the exam for the same module were $60.39\% \pm 17.61\%$ and $65.56\% \pm 18.89\%$, for the two consecutive semesters which was not statistically significant ($P=0.2$). In module 2, the multiple-choice questionnaire consisted of 37 and 31 questions in the first and the second semesters, respectively. The discrimination coefficients, in the first and the second semesters, were 0.31 ± 0.17 and 0.30 ± 0.18 ($P=0.72$), and difficulty coefficients were $63.7\% \pm 18.62\%$ and $67.7\% \pm 14.75\%$, respectively which none of them were statistically significant ($p=0.34$). The multiple-choice questionnaire in module 3 comprised 54 and 44 multiple-choice questions. The discrimination coefficients in module 3 were 0.33 ± 0.27 and 0.39 ± 0.27 ($P=0.42$) and their difficulty coefficients were $57.95\% \pm 19.22\%$ and $55.79\% \pm 22.88\%$ ($P=0.69$), respectively. As P-values showed, t-test determined that the differences were not significant.

The correlation coefficient of questions designed by two faculty members for periodontics module 1 in the first semester were 0.95 and 0.86 and for the second semester, it was 0.81 for all three faculty members. In module 2, the correlation coefficients for the first semester were 0.67, 0.66, and 0.94, while in the second semester; these coefficients were 0.63, 0.86 and 0.89 for each

Table 1. Difficulty coefficient of theoretical courses of Periodontology in first and second semester. Values are represented as mean±standard deviation

Periodontics	Difficulty Coefficient		P value
	First Semester	Second Semester	
module 1	60.39%±17.61	65.56%±18.89	0.2
module 2	63.7%±18.62	67.7%±14.75	0.34
module 3	57.95%±19.22	55.79%±22.88	0.69

Table 2. discrimination coefficient of theoretical courses of Periodontology in first and second semester. Values are represented as mean±standard deviation

Periodontics	Discrimination Coefficient		P value
	First Semester	Second Semester	
module 1	0.38±0.18	0.29±0.18	0.017*
module 2	0.31±0.17	0.30±0.18	0.72
module 3	0.33±0.27	0.39±0.27	0.42

*= significant

faculty member. The correlation coefficient for each faculty member in the first semester were 0.84, 0.79, and 0.84, and in the next academic term, they were 0.90, 0.83, and 0.82, for module 3. The mean compliance percentage to Millman checklist are presented in (Table 3). The mean compliance percentage in module 1 for the first and second semesters was %87.70 and %91.87, respectively (P=0.57). In the module 2, the mean percentages were %88.18 and %93.3 (P=0.17), whereas, in module 3, it was %80.69 and %94.48 for the first and second semesters, respectively (P=0.02).

Discussion

According to the findings of the present study, using the educational pamphlet was effective to improve the quality of multiple choice questions designed by the faculty members. It seems that in addition to evaluating the ability to remember specific facts and demonstration of technical skills, students should be assessed for their capacity to interpret data and to analyze facts and data within a given context and use it in unique situations, which require critical thinking and problem solving. The aim of the evaluation in educating health professionals is to verify each student's

capability to combine and employ the different areas of learning. Since these evaluations could mutually characterize students' capability to practice, they should take place frequently and over an extended period of time, in a work environment that simulates the actual work settings where health care providers interact with patients [2].

Comprehensive care stipulates placing the patient at the center of considerations and cares for the entire patient and all his or her needs and requirements, not just the medical and physical ones. As a result, a student would be accountable for the complete examination, treatment planning, supplying and providing the treatment, and reviewing and revising the treatment plan for and with a patient. Comprehensive care is a key feature of "competency profiling" which should support the assimilation and unification of all disciplines to the benefit of dental students and also patients who are receiving treatment. "Profile" means an acknowledged dentist who is a specific professional quite separate from a medical doctor. "Competencies" denote a series of general and specific knowledge, skills, and behaviors that qualify graduating dentists to apply and promote proper oral health management strategies in

Table 3. Results of questionnaire evaluation based on the Mean Millman checklist indicators for the theoretical modules of Periodontology (values are in percentage)

Periodontics module	Semester	completeness of stem information	Specific objective for question	clear vocabulary for stem	clear words for options	lack of negative option for the negative stem	lack of "none of above" and "all items" options	lack of contrasting option	Positive words in the stem	Independence of question	Compliance with the length and style of writing the options	Correct spelling of the stem	Correct spelling of the options	No duplicate options	Vertical options
Module 1	1	70	70	90	100	100	100	100	90	100	90	100	100	100	20
	2	75	84	87	89	100	100	100	93	100	77	96	93	87	100
Module 2	1	59	72	94	89	91	94	97	70	89	100	97	100	100	59
	2	65	94	100	90	100	100	100	94	100	74	97	100	94	100
Module 3	1	48	96	75	96	100	77	100	88	96	75	77	90	77	27
	2	75	97	86	95	97	90	100	95	100	88	100	100	95	100

order to improve the general health of their patients [6].

Multiple choice questions are able to evaluate and assess knowledge, understanding, and analyzing the power of students. Valuable and beneficial MCQs can be identified by analyzing items especially based on difficulty index [2, 8].

The difficulty coefficient (P) or as it might also call "ease index", is the percentage of individuals who have answered a question correctly and it ranges between 0 and 100%. The higher the coefficient, the easier the question, and vice versa, the coefficient closer to zero, means more difficult question. A question would be considered difficult if $P \leq 25\%$; questions with the P value between 26-75% would weigh up as medium difficult; if values between 76-95%, questions considered medium- low difficult; and it would be easy in the case of $P > 95\%$ [10]. In general, the difficulty coefficient of 30% to 50% gives the maximum amount of information about the differences between students [3].

Findings of the present study on the exam questionnaires of the theoretical portion of periodontics module in one academic year revealed that the majority of questions had a mean difficulty

coefficients of 60-67% which classified them as medium difficult and it is acceptable for general dentistry students.

Discrimination index (D) measures the power of distinction of a question in order to distinguish between strong and weak examinees. In other words, it specifies to what extent the question is able to separate the strong student from the weak. Discrimination index range lies between -1 to +1 and the higher index value means that the question was able to separate and identify the higher performing students from the lower ones, more precisely. The discrimination power of a question would be considered "very good" if D is more than 0.5; questions with the discrimination index of 0.3-0.5 would be graded as "good". While if it is between 0.2-0.3, the discrimination power would be categorized as moderate and D lower than 0.19 means that the discrimination power is weak [3,11]. However, according to a number of studies, 0.15 would be the cut-off point [2,8].

When the value of D is negative ($D < 0$), it simply means that students with the lower score answer more correctly to that particular question compare to those with the higher score. These undesirable situations could happen due to complex nature of

item which makes it possible for the least knowledgeable students to select correct response without any real understanding [2,8].

In the present study, discrimination coefficient varied between 0.2 and 0.4 in both semesters, which means the discrimination index of the questions generally lied between moderate to good. Statistical analysis revealed that the discrimination index of the first semester exam in the Periodontics module 1 was more superior compared to the second semester ($P=0.017$).

The coefficient of agreement of a question is defined by the correlation coefficient of each faculty member questions with the total scores of the exam. Higher coefficient means that the question is more compatible with the other questions in the questionnaire and has been able to serve its educational purposes [3]. The average coefficient of agreement of two consecutive semesters in the present study was 0.8 which can be considered desirable. In the present study questions of Periodontics modules were evaluated and compared in two consecutive semesters (one year), and on average, a coefficient of 0.8 was obtained.

According to Millman checklist, compliance with the structural principles of questions of periodontics modules in the first semester was between 80 and 88 percent, however, this value was increased to 92-94 percent in the second semester after professors and lecturers have studied the educational pamphlet. In this respect, compliance with the structural principles in Periodontics reached a favorable amount. Significant improvement happened at the second semester in Millman checklist indicators such as compliance, without using the options of "none of above" and "all items", completeness of the information of stem, and verticality of the options. Overall, mean values of Millman checklist indicators of the second semester was higher than the first semester, which implies that faculty members have considered these issues in designing the questions. This might be due to the positive impact of the educational pamphlet on the design of the final exam through draw faculty members' attention to the points raised in the pamphlet.

The quality of questions and student's assessment are very important since students are generally

encouraged and driven by it. Well-designed questions can identify the talented and diligent students from less encouraged and underachiever students. Therefore, educators can motivate and inspire more the strong student and employ other methods of education in order to improve and assist low scoring students in achieving their highest potentials. Refinement and enhancement in the quality of MCQs will enhance and increase the validity of the examination as well as students' deep learning attitudes. Questions with high P (easier questions) discriminate poorly; while questions with a low P (difficult questions) are good discriminators, except when they are so difficult, even good students cannot be able to answer them correctly [8].

A properly assembled and outlined MCQ should be examined for the standard or quality. Item analysis is a valuable tool and a relatively simple procedure which can be performed after the examination in order to provide information concerning the reliability and validity of the test. It also considerably helps to improve the quality of items and prepares a viable question bank for succeeding use. Furthermore, it is beneficial to both students and teachers since it provides feedback to the teacher to enhance their method of teaching and encourage the learners to learn more effectively.

It would be better to conduct this assessment for several semesters, and each time the evaluation results should be presented to the faculty members, so these positive changes happen over the course of time.

In studies conducted about the effect of educational interventions on optimizing the design of multiple-choice questions, the results showed the positive effects of these training on the quality of the questions [8]. It seems that the positive changes of these coefficients in the present study were because of holding workshops and providing a compact pamphlet on different methods of teaching as well as on how to design questions, which caused an enhancement in the faculty members' performance. Although a favorable change in the indicators of Millman checklist was seen in the second semester, it was not statistically significant ($P=0.65$). The most positive changes were observed in indicators of "completeness of

the information of stem", "verticality of the options", and lack of "None of above" and "all items" options.

Examination and assessment could be a source of nervousness, stress, and anxiety for the undergraduate medical and dental students, nonetheless, in reality, it guides and gives a direction to study harder and improve students' skills. Hence, it is essential in addition to stimulate their cognitive skills during the teaching, their higher mental and reasoning skills should be examined frequently [1].

As part of the ongoing professional improvement in the dental schools, faculty members should have access to the various form of educational development programs such as workshops in which they would have opportunities to discuss learning, teaching, and assessment with other fellow educators. Experts also should be invited to guide and conduct the dialogue toward particular important issues. A comfortable and relaxed setting gives faculty members the chance to participate and share their own knowledge, experience, and

opinions with other colleagues. Therefore, the faculty members would have information and motivation to recreate and get excellent in their teaching techniques [6]. To achieve the best level of efficiency and audience attraction, the use of modern training techniques and the role of qualified lecturers should not be neglected.

Conclusion

According to the findings of the present study, using the educational pamphlet did not significantly improve the quality of multiple-choice questions designed by the faculty members, in spite of slight improvements in some Millman checklist items. Therefore, it seems that holding short courses and workshops would be beneficial to improve faculty members' performance in designing questions which could lead to a better distinguish between diligent and less encouraged students. Keep in mind that the proper place and time for holding these workshops is of utmost importance. Workshops with the shorter length which held at the workplace could have better efficiency.

Acknowledgment

This work was supported by Dental School of Tehran Azad University of Medical Sciences.

References

1. Khan MU, Aljarallah BM. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for assessing the cognitive skills of undergraduate medical students. *Int J Health Sci (Qassim)*. 2011 Jan; 5 (1):39-43.
2. Gajjar S, Sharma R, Kumar P, Rana M. Item and Test Analysis to Identify Quality of Multiple Choice Questions (MCQs) from an Assessment of Medical Students of Ahmedabad, Gujarat. *Indian J Community Med*. 2014 Jan;39(1):17-20.
3. Taib F, Bahri Yusoff MS. Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences*. 2014 June; 9(2):110-114.
4. Newell FW, Leopold IH, Adler FH. The multiple-choice question test of the American Board of Ophthalmology. *Trans Am Ophthalmol Soc*. 1970;68:165-85.
5. Mehrens WA, Lehmann IJ. Measurement and evaluation in education and psychology. 4th ed. New York, Wads Worth publishing, 1991:145-162.
6. Nulty DD, Short LM, Johnson NW. Improving Assessment in Dental Education through a Paradigm of Comprehensive Care: A Case Report. *J Dent Educ*. 2010 Dec;74(12):1367-79.
7. Wilkerson L, Irby DM. Strategies for improving teaching practices: a comprehensive approach to faculty development. *Acad Med*. 1998 Apr; 73(4): 387-96.
8. Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK, Syed S, et al. Faculty development programs improve the quality of Multiple Choice Questions items' writing. *Sci Rep*. 2015 Apr 1;5:9556.
9. Albino JE1, Young SK, Neumann LM, Kramer GA, Andrieu SC, Henson L, et al. Assessing dental students' competence: best practice recommendations in the performance assessment literature and investigation of current practices in predoctoral dental education. *J Dent Educ*. 2008

Dec;72(12):1405-35.

10. Urrutia J, Orrego M, Wright AC, Amenabar D. An assessment of the Chilean national examination of orthopaedic surgery. BMC Med Educ. 2016 Mar

1;16:78.

11. Crocker L, Algina J. Introduction to classical and modern test theory. 4th ed, USA: Cengage learning, 2008: 313-20.