Evaluation of Multiple Choice Questions of Oral and Maxillofacial Medicine Courses 1, 2, and 3 in the First Semester of Academic Year 2014-2015

M Rezai1, T Farrokhnia2*, M Vatanpour1, S Lesan2, Sh yazdipour4

1Associate professor, Oral Medicine Dept, Dental Branch of Tehran, Islamic Azad University, Tehran, Iran.
2Assistant professor, Oral Medicine Dept, Dental Branch of Tehran, Islamic Azad University, Tehran, Iran.
3Assistant professor, Endodontics Dept, Dental Branch of Tehran, Islamic Azad University, Tehran, Iran.
4Dentist

ABSTRACT

Background and aim: Lack of proper assessment of students’ knowledge regarding theoretical contents leads to disappointment and negligence by weak students. This study assessed multiple-choice questions (MCQs) of oral and maxillofacial medicine courses in the first semester of the academic year 2014-2015.

Materials and methods: In this cross-sectional study, questions of theoretical oral medicine courses 1, 2, and 3 were assessed according to twelve items in Millman’s checklist, including clearness of stem, negative option for stem, specific option, contrastive option, positive words in stem and options, writing structure of stem, duplicated option, the spelling of stem and options, vertically of options, positivity of stem and options, and use of “all of the above” and “none of the above” phrases in options. Difficulty and discrimination coefficients of each question and the rate of compliance with Millman’s principles were evaluated using Excel 2007 and SPSS 16. Data were analyzed using descriptive statistics, analysis of variance (ANOVA), and Regression analysis.

Results: The responses of 219 students to 113 MCQs were evaluated. In theoretical course 1 with 40 MCQs, discrimination coefficient was 0.36±0.2, difficulty coefficient was 73.2±15.02, and correlation coefficient was 0.58. In theoretical course 2 with 35 MCQs, discrimination coefficient was 0.2±0.04, difficulty coefficient was 70.91±25.13, and correlation coefficient was 0.47. In theoretical course 3 with 38 MCQs, discrimination coefficient was 0.3±0.19, difficulty coefficient was 68.73±26.89, and correlation coefficient was 0.5. The percentage of compliance with Millman’s principles was 81.88%, 83.58%, and 84.42% in courses 1, 2, and 3, respectively.

Conclusion: The weakest discrimination coefficient was noted in theoretical oral medicine 2. The high percentage of simple difficulty coefficient in the three theoretical courses indicates the necessity of training faculty members in designing questions.

Introduction:

One of the concerns of the academic community, including students and faculty members, is the issue of assessing students with regard to the degree of learning the theoretical educational contents.\(^{(1)}\) This concern is raised not only with regard to the educational contents in the entire semester but also regarding every lesson and even for all the questions related to one lesson\(^{(2)}\). For the first time, Fredrick Kelly posed the topic of the quality of multiple-choice questions (MCQs) in 1914.\(^{(3)}\) Nowadays, faculty members of different departments try to implement the best questions, the highest number of questions, and the highest number of options for a better assessment of students.\(^{(4)}\)

One of the methods for preparation of theoretical questions is MCQs, which Millman has provided a checklist for evaluating their design accuracy.\(^{(3)}\) Due to its nature and particular complexities, evaluation is one of the most extensive and controversial educational topics, and if properly executed, the results can identify the strengths and weaknesses of education.\(^{(5)}\) Without a doubt, the type and the quality of exams directly affect the teaching method and the teacher’s credibility. Therefore, it is appropriate to prepare and execute tests with precision and integrity so that the test can be implemented as a reliable and valid tool for assessing the learners who are the future-makers in different fields in the country.\(^{(6)}\)

MCQs were first invented by Fredrick Kelly in 1914.\(^{(3)}\) This type of questions was first examined five decades previously in written board certification exams in the United States.\(^{(7)}\)

The evaluation of difficulty level and discrimination power was performed by test-developing specialists, Mehrens and Lehmann in 1984.\(^{(8)}\) In some previous studies, it was concluded that the design of the end-of-the-semester tests does not have a proper and satisfactory status, and generally, the tests used in the process of evaluating the educational outcomes of students have fundamental structural flaws and do not act as an accurate and valid tool. Also, they do not measure the actual differences between the examinees.\(^{(9)}\) Given the lack of information on the status of questions in medical universities and oral and maxillofacial medicine departments, we aimed to assess the quality of questions of theoretical oral and maxillofacial medicine courses, according to Millman’s standard, and their difficulty and discrimination coefficients.

Materials and Methods:

In this cross-sectional research, the questions of theoretical oral and maxillofacial medicine courses 1, 2, and 3 in the first semester of the academic year 2014-2015 were studied. The questions were evaluated in accordance with twelve items in Millman’s checklist, including stem clearness, negative option for stem, specific option, contrastive option, positive words in stem and options, writing structure of stem, duplicated option, the spelling of stem and options, vertical writing of options, positivity of stem and options, the use of “all of the above” phrase, and the use of “none of the above” phrase in options. Through examining the question type, it was assessed that whether Millman’s principles were met\(^{(3-13)}\) and the score of each question as well as the total score of the questions in that semester were determined. The papers were examined with students’ answer sheets, and the score of each question of oral and maxillofacial medicine courses and the total score were determined and recorded with student identification (ID).

The difficulty and discrimination coefficients were determined for each question; the difficulty coefficient is the percentage of the total number of people who answered a question correctly divided by the number of the examinees,\(^{(3)}\) and the discrimination coefficient is the correct choices of the upper group minus the correct choices of the lower group divided by the number of people in a group (upper or lower).\(^{(3)}\)

At the end of the research, the difficulty coefficient of each question, the discrimination coefficient of each question, the correlation of the total score with the answer to each question, and the percentage of answers for each question from the available choices were determined and were calculated and statistically analyzed by Excel 2007 and SPSS 16 software programs, using descriptive
statistics, analysis of variance (ANOVA), and Regression analysis.

**Result:**
The responses of 219 students to 113 MCQs of theoretical oral and maxillofacial medicine courses 1, 2, and 3 were evaluated. (table 1)

In theoretical oral and maxillofacial medicine course 1 with 40 MCQs and 73 students, the discrimination coefficient was 0.36±0.2, the difficulty coefficient was 73.2±15.02, and the correlation coefficient was 0.58. In theoretical oral and maxillofacial medicine course 2 with 35 MCQs and 61 students, the discrimination coefficient was 0.2±0.04, the difficulty coefficient was 70.91±25.13, and the correlation coefficient was 0.47. In theoretical oral and maxillofacial medicine course 3 with 38 MCQs and 85 students, the discrimination coefficient was 0.3±0.19, the difficulty coefficient was 68.73±26.89, and the correlation coefficient was 0.5. Regarding the correlation coefficient of the questions, it should be noted that conformity was observed between the response and the total score only in questions 6, 8, 9, 11, 13, 17, 21, 25, 26, 34, 35, and 40 in theoretical oral and maxillofacial medicine course 1, in questions 8, 30, 32, and 33 in theoretical oral and maxillofacial medicine course 2, and in questions 2, 8, 10, 15, 26, 29, 31, 37, and 38 in theoretical oral and maxillofacial medicine course 3.

According to the difficulty coefficient, in theoretical oral and maxillofacial medicine course 1, of 40 MCQs, 72.5% were simple, 25% were moderately difficult, and 2.5% were difficult. (Table 2)

<table>
<thead>
<tr>
<th>Oral and maxillofacial medicine course</th>
<th>Number of questions</th>
<th>Number of students</th>
<th>Discrimination coefficient</th>
<th>Difficulty coefficient</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>73</td>
<td>0.36±0.2</td>
<td>73.2±15.02</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>61</td>
<td>0.2±0.04</td>
<td>70.91±25.13</td>
<td>0.47</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>85</td>
<td>0.3±0.19</td>
<td>68.73±26.89</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 2: Comparison of difficulty coefficients of questions divided by oral and maxillofacial medicine courses

<table>
<thead>
<tr>
<th>Oral and maxillofacial medicine course</th>
<th>Difficult</th>
<th>Moderately difficult</th>
<th>Simple</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.5%</td>
<td>25%</td>
<td>72.5%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>37%</td>
<td>63%</td>
</tr>
<tr>
<td>3</td>
<td>13%</td>
<td>26%</td>
<td>61%</td>
</tr>
</tbody>
</table>

In theoretical oral and maxillofacial medicine course 2, of 35 MCQs, 63% were simple, 37% were moderately difficult, and 0% were difficult. In theoretical oral and maxillofacial medicine course 3, of 38 MCQs, 61% were simple, 26% were moderately difficult, and 13% were difficult.

Of the twelve Millman’s principles, the indices of non-repeatability of options, the spelling of stem and options, and positivity of stem and options had the highest percentage of compliance. (Table 3)

The lowest percentage of compliance with Millman’s principles was associated with the index of verticality of options.

The highest percentages of compliance with Millman’s principles were as follows:
In theoretical course 1: non-repeatability of options, the spelling of stem and options, positivity of stem and options, and no use of the “none of the above” phrase in options.

In theoretical course 2: writing structure of stem, non-repeatability of options,
Evaluation of Multiple Choice Questions of Oral and Maxillofacial Medicine Courses

Table 3: Compliance with each of Millman's principles according to indices and divided by oral and maxillofacial medicine courses

<table>
<thead>
<tr>
<th>Oral and maxillofacial medicine course</th>
<th>Clarity of stem</th>
<th>Lack of negative option for stem</th>
<th>Specific option</th>
<th>Lack of contrastive option</th>
<th>Positive words in stem</th>
<th>Writing structure of stem</th>
<th>Lack of duplicated option</th>
<th>Spelling of stem and options</th>
<th>Vertically of options</th>
<th>Positivity of stem and options</th>
<th>No use of &quot;all of the above&quot; phrase in options</th>
<th>No use of &quot;None of the above&quot; phrase in options</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>31</td>
<td>30</td>
<td>35</td>
<td>32</td>
<td>38</td>
<td>40</td>
<td>40</td>
<td>2</td>
<td>40</td>
<td>38</td>
<td>40</td>
<td>87</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>31</td>
<td>20</td>
<td>34</td>
<td>32</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>1</td>
<td>35</td>
<td>33</td>
<td>35</td>
<td>69</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>35</td>
<td>30</td>
<td>36</td>
<td>33</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>4</td>
<td>38</td>
<td>32</td>
<td>37</td>
<td>71</td>
</tr>
</tbody>
</table>

Discrimination coefficient, denoted by the letter D, indicates the power of the question in separating the strong group (who gained high scores in the test) from the weak group (who gained low scores in the test) and can have values between -1 and +1; larger coefficients indicate better results. In its interpretation, if D ≥ 0.5, the discrimination power of the question is very good, 0.3 ≥ D ≥ 0.2 denotes a moderate power, while discrimination coefficients of 0.19 ≥ D are considered weak to very weak. (3) The discrimination coefficients of theoretical courses 1 and 3 were moderate to good, and the discrimination coefficient of theoretical course 2 was weak to moderate. Theoretical course 2 was inferior, with 8% of its questions having a discrimination coefficient of zero, that is, they have by no means been able to distinguish between students of the strong group and the weak group, and 5.7% of the questions have had a negative discrimination coefficient, which means that students from the stronger group have had a worse performance than the weaker group; substantial revision of such questions is recommended. The reason for the presence of such coefficients can be that the students of the strong group have not fully understood the question or that they have misunderstood it.

Another important finding in this section was the difficulty coefficient that is represented by the letter P, which is the percentage of people who have answered a question correctly, and it contains values between 0 and 1. In its interpretation, it should be stated that P < 50% represents a difficult question, 75% ≥ P ≥ 50% represents a moderately difficult question, and P > 75% denotes a simple question. (3) The questions in oral and maxillofacial medicine course 1, based on the difficulty coefficient, were 72.5% simple, 25% moderately difficult, and 2.5% difficult. In theoretical course 2, 63% of the questions were simple, 37% were moderately difficult, and 0% were difficult.
In theoretical course 3, 61% of the questions were simple, 26% were moderately difficult, and were 13% difficult.

Another important and valuable finding of this research was the correlation coefficient between the total score and the answer to each question; the higher the correlation coefficient, the more the question is matched with the set of other questions, thus differentiating between the strong and weak examinees with a higher precision, and the lower the coefficient, the lower the conformity. The correlation between the total score and the response to each question in theoretical oral and maxillofacial medicine course 1 was 0.58%, while in theoretical course 2, it was 0.48%, and in theoretical course 3, it was 0.5%; which indicates a moderate correlation coefficient between the questions and the whole test.

In our study, the rate of ineffective options in theoretical oral and maxillofacial medicine course 1 was 4.3%, while in theoretical course 2, it was 12.14%, and in theoretical course 3, it was 6.5%, which indicates that the faculty members of the oral and maxillofacial medicine department may need more precision in eliminating such options. The rate of ineffective options in a study by Sayar et al was 2.9% in endodontics course 1, 3.9% in endodontics course 2, and 3.49% in endodontics course 3, indicating fewer ineffective options compared to our study.

The present study showed that the percentage of compliance with Millman’s principles was 81.88% in theoretical oral and maxillofacial medicine course 1, 83.58% in theoretical course 2, and 84.42% in theoretical course 3. With a very insignificant difference, the lowest percentage was related to the questions of theoretical oral and maxillofacial medicine course 1, and the highest percentage was related to theoretical oral and maxillofacial medicine course 3 such that the least amount of compliance was in the verticality of options, and the maximum amount was related to the spelling of stem and options, writing structure of stem, positivity of stem and options, and non-repeatability of options.

Hosseini Teshnizi et al analyzed the MCQs of non-continuous undergraduate medical records in 2010. To this end, all MCQ tests at the end of the semester related to non-continuous undergraduate medical records at Hormozgan University of Medical Sciences were selected. Then, the difficulty and discrimination indices of the questions were analyzed using Excel 2007 and SPSS 16 software programs by descriptive statistics, ANOVA, and Regression analysis. The results showed that from 128 questions, 105 (57.7%) had a suitable difficulty coefficient, the least of which was associated with research methodology (59%), and the highest was related to medical terminology (78%). The discrimination power of 133 questions (73.1%) was weak to very weak such that the English language (0.05) and common classification systems (-0.03) had the weakest discrimination power, and research methodology (0.23) and computer applications (0.18) had the highest discrimination power. The results of statistical analyses showed that there was a significant difference in the difficulty level and discrimination power in all the courses. Also, there was no relationship between the difficulty level and discrimination coefficients of the questions. Although the difficulty coefficient for all tests had an acceptable and appropriate level, the results of the discrimination index showed that for almost all quiz questions, the quality was weak to very weak. Therefore, it is better to remove these questions from the question bank or to fundamentally revise them.

One of the factors indicating the higher accuracy of our study is the categorization of difficulty coefficients to simple, moderately difficult, and difficult, which is not observed in the cited study. Discrimination coefficients of theoretical oral and maxillofacial medicine courses were higher in the present study, indicating a greater distinction between the strong and weak groups in our study. Unlike our study, Millman’s checklist has not been used to evaluate the quantity of questions in the abovementioned study.

Meyari and Beiglarkhani assessed the effect of educational intervention on optimizing the design of MCQs of annual residency exams at the School of Dentistry of Hamadan University of Medical Sciences in 2008-2009. In the cited quasi-experimental study, all of the questions of annual residency exams at the School of Dentistry, Hamadan
Evaluation of Multiple Choice Questions of Oral and Maxillofacial Medicine Courses

University of Medical Sciences in 2008 were evaluated with regard to the design and structure using a valid and reliable checklist based on Millman’s principles. After conducting a workshop for test developers in both academic years, the questions of the residency exam of the year 2009 were also examined. The data before and after the intervention related to all the questions were compared using ratio comparison tests. Out of 1239 questions in the mentioned study, 63.1% and 76.3% of the questions, related to the years 2008 and 2009, respectively, were designed without any flaw, and the difference was statistically significant. The results of the cited research with regard to compliance with Millman’s structural principles compared to our study (compliance with Millman’s structural principles: 81.88% in theoretical oral and maxillofacial medicine course 1, 83.58% in theoretical course 2, and 84.42% in theoretical course 3) are moderate and lower (63.1%).

Haghshenas et al examined MCQ tests at Mazandaran University of Medical Sciences in the first semester of the academic year 2006-2007. In this descriptive study, all the questions of written MCQ tests in one semester at Mazandaran University of Medical Sciences were reviewed. The questions were evaluated with regard to the structure using Millman’s checklist for stem and options design. Out of 1478 questions in 25 tests, the percentages of questions related to basic sciences, physiopathology, and internship were 28.7%, 18.7%, and 52.6%, respectively. In terms of structural errors, 46% of the questions were flawless, and the rest had one or several structural errors. Taxonomy and structural errors of internship and physiopathology tests were significantly lower compared to tests of basic sciences.

In the mentioned research, questions without structural errors were 46%, while in our research, this figure was 95% in oral and maxillofacial medicine course 1, and 100% in oral and maxillofacial medicine courses 2 and 3. It is possible that the faculty members of the oral and maxillofacial medicine department may have completed training sessions and courses regarding the design of standard questions. Unlike our study, in the cited study, the difficulty and discrimination coefficients of MCQs have not been calculated.

Roshanpour et al evaluated the effect of the psychiatry questions. Our study has a higher accuracy compared to the cited study. Categorization of difficulty coefficients to simple, moderately difficult, and difficult was not observed in the mentioned study. The magnitude of difficulty coefficients was higher in our study. In addition, Millman’s checklist was not used to check the quantity of questions in the cited study.

Although the use of educational pamphlets did not significantly improve the quality of MCQs in the study by Sayar et al, some of the items in Millman’s checklist showed some improvements that justify the use of educational pamphlets to some extent.

Conclusion

The results of the present study showed that the weakest discrimination coefficient was noted in theoretical oral and maxillofacial medicine course 2. The simple difficulty coefficient in the three theoretical courses showed a high percentage, which requires planning by the Education Development Centers of universities to train and empower faculty members in designing questions, especially in the mentioned fields. Due to the importance of evaluations in dentistry, it is recommended to conduct this study in other departments as well.

Aknowledgement:

We are thankful to those who helped us in this article including research center of Islamic Azad university –Dental Branch.

References:
1. Hosseini Tesnizi S, Zare SH, Solati SM. Quality analysis of multiple choice questions (MCQs) examinations of noncontinuous undergraduate medical records. hmj 2010, 14(3): 177-83
4. Shakurnia A, Khosravi A, Shariati A, Zareei A; Survey on multiple choice questions of academic members in Ahvaz Jundishapur university
promotion program of faculty members and the results of medical promotion exams at Ker-
manshah Health and Education University.\(^\text{(12)}\)

In a quasi-experiment, 105 question designers of medical residency promotion exams were selected by census sampling method, and the results of the tests in 2008 and standard test design pamphlets were sent to them. Quantitative data collection tool included Ministry of Health test analysis report on discrimination and difficulty coefficients, the standard deviation of questions, and test reliability, and for qualitative test data, the tool includes the report of the Medical Education Secretariat on the percentage of questions without structural errors, paired and independent. The data were analyzed by ANOVA. The mean discrimination coefficient, the validity of the test, and the percentage of questions without structural errors were determined. The lowest difficulty coefficient (0.54) was observed with urology questions, while the highest difficulty coefficient (0.67) and the highest percentage of increase in the discrimination coefficient were detected in of medical sciences in 2006. Strides in development of Medical Education (Special supplement for 8th National congress of medical education) Kerman, 2007:44.


10. Meyari A, Beiglarkhani M. Improvement of Design of Multiple Choice Questions in Annual...