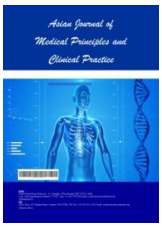

Asian Journal of Medical Principles and Clinical Practice

3(2): 26-34, 2020; Article no.AJMPCP.54264

The Quality of MCQs in Radiology Theory Lessons Assessed by the Millman Standard

Nazanin Forghani[1*]

[1]*Department of Dental Surgery, Islamic Azad University, Iran.*

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

Editor(s):
(1) Dr. Karthik Yadav Janga, Bayer Healthcare, USA.
(2) Dr. Suprakash Chaudhury, Dr. D. Y. Patil Medical College, Hospital & Research Centre, India.
Reviewers:
(1) Kavita Nagar, Rajiv Gandhi University of Health Sciences, India.
(2) Sameeulla Shaik, Dr. NTR University of Health Sciences, India.
(3) Michael Bordonaro, Geisinger Commonwealth School of Medicine, USA.
Complete Peer review History: http://www.sdiarticle4.com/review-history/54264

Original Research Article

Received 28 December 2019
Accepted 04 March 2020
Published 24 March 2020

ABSTRACT

Background: One of the real concerns of society, especially students and teachers, in particular, is the issue of student assessment in content learning theory. If the assessment of these questions cannot asses students' knowledge, it discourages active students, and, on the other hand, less attention leads to having poor students, which has known implications, especially in dentistry that endangers the health of the community.
Aim: Therefore, this research was conducted to evaluate the radiological multiple-choice questions 1, 2 and 3 in the second semester of the academic year 2016-2017.
Methods: This study was a Cross-Sectional one and data were calculated by ANOVA and regression analysis.
Results: During the study, 278 students' answers of 611 questions were evaluated in theoretical radiological lessons 1, 2 and 3. On radiology 1 out of 40 multiple choice questions and 102 students, discrimination indexes were 0.23 ± 0.15 and difficulty coefficient % 54.8 ± 92.43 and the correlation coefficient was 0.64 and on Radio 2 from the 36 multiple-choice question discrimination index was 0.81 ± 0.61, difficulty coefficient % 37.86 ± 21.69 and the correlation were 0.48. On Radio 3 out of 40 multiple-choice questions, the discrimination coefficient was 0.40 ± 0.02, correlation coefficient % 65.37 ± 17.89 and the correlation was 0.68. In examining ways of Millman.

**Corresponding author: E-mail: nazfor86@gmail.com;*

These principles were met in radiology 1 (% 96.75 ± 4.50) and in radiology 2 (% 95.65 ± 5.58) and in radiology 3 (% 29/50 ± 8/12).
Conclusion: It shows that radiology questions 1 and 2 were low discrimination index and difficulty coefficient Radio 1, 2 and 3 was average.

Keywords: Millman; discrimination index; difficulty index; radiology.

1. INTRODUCTION

The most important concerns of society, especially students and teachers is the issue of student assessment in content learning theory. This concern has not only been in every year in all terms but also every lesson even for all questions of a lesson [1].

Currently, professors in the department try to apply the best questions and options for a better assessment of students [2]. But if all these measures are not effective and the evaluation of these questions cannot asses the students ' knowledge, it discourages students from studying and activating, and on the other hand, less attention results in having poor students, which has known implications, especially in dentistry which endangers the health of the community [3]. One of the ways to write theoretical questions is Multi-Choice Question (MCQ), and there are several ways to design these questions. One of the best practices is Millman method [4].

Multiple-choice questions were first introduced by Fredrick in 1914 (3). These types of questions were first examined five decades ago in the written examinations of the specialized boards of the United States [5].

1) The main part of the trunk of the question: The main text of the question consists of the question and the question is to be measured by the question.
2) The correct option or the answer to the question: it is one of the suggested options which should be selected by the examiner, this is called the key.
3) Deviant options: Apart from the correct option, the number of other options are also provided for each question, which is referred to as deviating options. The original text is a phrasing sentence or an unfinished sentence, and the suggested options are incomplete, such as the use of the sentence or completeness.

- Investigating the observance of structural principles in multiple-choice questions related to the question of stems and options by Millman in 1981 [4].

There are some tips to observe structural principles in multiple-choice questions that are explained in the following:

1) Each question should measure an important subject or educational goal.
2) The questions must be written clearly and use phrases and words to be understood.
3) Do not repeat the content in the options.
4) The main points of the questions should be completely written in the question body.
5) The options for a question must be consistent and relevant to the subject matter.
6) The question is to be written that the correct answer is just the right answer or the correct answer.
7) The diversion options should be written in such a way as to attract the attention of the untrained testers to the question.
8) The questions should be written in such a way as to be correctly and grammatically supplement of the question.
9) Highlight negative words in negative questions.
10) Avoid writing questions in which text is negative and the options are negative.

Write questions

11) As far as possible, use of "all of the above," "none of the above," "none of them" should be avoided.
12) Avoid Misleading Questions.
13) The opposite alternative, one of which is correct, is not used.
14) In the questions where their sentence is incomplete, leave blank in the last part of the sentence.

A review of the difficulty and power of the test was carried out by Mehrens' specialist in designing questions [6].

Calculating the difficulty factor of the question: By definition, it is said to the percentage of the total number of examiners

who answer a question correctly is the difficulty index, shown by the letter "P". If all questions are involved in the analysis of a question, all the test sheets are sufficient to calculate the difficulty coefficient of the question based on this formula: the total number of people who choose the correct answer divided by the total number of the examiners and the result multiplied by 100.

Calculation of the discrimination coefficient of the question: The power of the question in the differentiation or recognition between the strong and weak members of the circle, which indicates how much the question can separate the strong group from the weak group and display by letter D.

Haghshenas et al. [7] studied the exams of Mazandaran University of Medical Sciences [8] and concluded that the quality of multiple-choice tests in terms of taxonomy distribution and observance of the principles of MCQ structure was different among different section tests and in some cases require reformative action.

Vahid Shahi et al. [9] in the study of taxonomy observed the structural principles in the questions and their differentiation advantage in the encyclical test before the educational interventions, have positively influenced the program of improving the quality of the Encyclopedia Test and the significant improvement of taxonomy and observing certain structural principles in the questions. Although the need for extension of these programs was demonstrated at this level and other levels of medical education [10].

Pourmirza Kalhori [11] studied the effect of promoting knowledge of the faculty members and the results of the medical assistant promotion tests of Kermanshah University of Medical Sciences and found that the average discrimination index, test validity, the percentage of questions without structural forms and the percentage of questions with taxonomy 2 and 3 increased significantly. The lowest coefficient of difficulty was urological questions and the highest difficulty factor and the highest percentage increase in the coefficient of differentiation in psychiatric questions.

The results indicate that the knowledge of designers has been increased by using educational pamphlets and provide feedback on the results of tests of past periods.

The question now is that just how much per cent of these questions are rich in content. There is a contradiction in the history of the question design [12]. About this gap and the lack of information about its status (quality of questions) in in the university and the radiology section in this research, we determined the quality of the questions of Radiology 1, 2 and 3 theoretical lessons from Millman's standard view.

2. MATERIALS AND METHODS

The research was carried out in a cross-sectional study. Radio questions 1, 2 and 3 were reviewed in the second half of the year 2016-2017. The questions were evaluated by Millman's 12 indicators. Twelve Millman indicators include stem transparency, negative selection of stems, selectable options, antonyms option, positive words in stems and alternatives, stem writing structure, option duplication, stem options and options, vertical write options, stem positive and options. The use of the expressions of all the cases and the use of the phrase in none of the options were evaluated by studying the type of question and from Seif's book, it was examined whether Millman's principles were observed or not? In this way, the score of each question and the total score of the questions during that term were allocated. It was evaluated by the students' responses and the score of each question was determined by the radio module and their total score and recorded with the student code. The difficulty and discrimination coefficient were determined for each question so that the difficulty index as the percentage of the number of people who choose the correct answer is divided by the number of examiners [4] and the discrimination index as the correct choices of the upper group minus the correct choices of the lower group divided by the number of people in a group (up or down) [4].

2.1 Statistical Society

All questions of radiology 1, 2 and 3 lessons in the second semester of the academic year 2016-2017 and students' responses.

2.2 The Statistical Sample

36 radiology lessons 1 in the second semester, 40 radiology lessons 2 in the second semester, 40 radiology lessons in the second semester. Total of three radiology courses, 116 questions and a total of 278 student responses were reviewed.

2.3 Sampling Method

In this research, sampling was done by the census method.

2.4 Statistical Analysis

At the end of the research, the hardness of each question, the discrimination index of the question, the correlation of the total score with the response to each question, the percentage of answers of each question were determined by the options and with the help of software 2007 Excel and 16Spss, using descriptive statistics, analysis of variance and regression analysis and analyzed statistically.

3. RESULTS

Throughout the survey, 278 students responded to 116 questions on radiology theoretical courses 1, 2 and 3. On radiology lesson 1, out of 40 multiple choice questions and 102 students, the discrimination index was 0.23 ± 0.15, the difficulty was 54.97 ± 29.49% and the correlation coefficient was 0.64. On radiology 2 out of 36 multi-choice questions and 85 students had the discrimination index of 0.18 ± 0.16, a difficulty factor of 37.86 ± 21.96 and a correlation coefficient of 0.48. On radiology 3, among 40 multiple choice questions and 19 students, the discrimination index was 0.40 ± 0.20, the coefficient of difficulty was 65.37 ± 17.89% and the correlation coefficient was 0.68. There were no significant differences between the three radio groups 1, 2 and 3 (P = 0.07).

On Radiology 1, from 40 multiple-choice questions, 5% of questions with very strong discrimination coefficient, 27.5% with good, and 22.5% with moderate, and 45% with weak discrimination had been seen. In Radio 2, from 36 questions, 8.3% with very strong coefficient, 13.8% with strong discrimination coefficient, 25% with moderate, 52.7% with the weak index had been observed. In Radio 3, from 40 multiple choices, 02% of the questions with very strong discrimination coefficient, 20% with good and 22.2% with moderate and 17.5% with weak coefficient had been seen.

On radiology 1, from 40 multiple-choice questions, 12.5% of the questions with high difficulty index and 35% with moderate difficulty and 52.5% with a low difficulty index had been seen. On Radiology 2, from 36 multiple-choice questions, there were 6.61% of the questions with high index, 63.8% with moderate and 19.4% with a low difficulty index. On radiology 3, from 40 multiple-choice questions, 55% of the questions had a high difficulty index and 27.5% and 17.5% of the questions had moderate and low difficulty factor respectively.

On radiology 1, the 20 questions which consisted of 50% of total questions are in 4 options, have a moderate correlation between the beta scores and their correlation coefficients of 0.44. On radiology [1], 12 questions, which consisted of 33.3% of the total questions included the correlated test beta-scores with a correlation coefficient of 0.48. In radio [4], 23 questions which consisted of 57.5% of the questions, had a moderate correlation with the total score of the test, which correlation coefficient is 0.68.

On radiology 1, from the total of questions correlated with the total test, 15% of the questions were with the strong correlation coefficient, 75% with moderate and 10% with the weak correlation coefficient. In Radio 2, from the total of questions correlated with the total test, 8.3% of the questions with strong correlation, 58.3% with moderate and 41.7% with low coefficients had been measured. On radiology 3, from the total of questions correlated with the total test, 16.2% of the questions with a strong coefficient, 65.2% with moderate and 8.7% with a low index had been seen.

In Millman Principles, these principles are on radiology 1 (96.75±4.50), of which 22 questionnaires contain 45% of the questions. Each of the 12 principles is fully (one hundred per cent). On radiology 2 (95.65 ± 5.58), of which two questions, which include 52.78% of the total questions, do not fully respect the twelve

Table 1. The rate of indicators of multi-choice testing by radiology classes

Radiology	Questions	Number of students	Discrimination index	Difficulty index	P
1	40	201	0.23 ± 0.15	54.97 ± 29.49	0.64
2	63	85	0.18 ± 0.16	37.86 ± 21.96	0.48
3	40	19	0.40 ± 0.20	65.37 ± 17.89	0.68

Table 2. Comparison of the rate of the coefficient of questions divided by radio lessons

Radiology	Very good discrimination %	Good discrimination%	Moderate discrimination %	Weak discrimination%
1	5	27.5	22.5	45
2	8.3	13.8	25	52.7
3	0.2	20	22.2	17.5

Table 3. Comparison of the difficulty coefficient of questions by radio curriculum

Radio curriculum	Low difficulty index %	Moderate difficulty index %	High difficulty index %
1	52.5	35	12.5
2	19.4	63.8	16.6
3	17.5	27.5	55

Table 4. Correlation of questions with total test and its value

Radiology lesson	Coherence coefficient	The percentage of coherent questions for the whole test
1	50	0.64
2	3.33	0.48
3	57.5	0.68

Table 5. Comparison of correlation coefficients of questions between radiology lessons

Radiology lesson	Strong coherence	Moderate coherence	Low coefficient
1	15	75	10
2	8.3	58.3	41.7
3	16.2	65.2	8.7

Table 6. Millman principles observing degree based on indicators and radio classes

Radio lesson	Full Millman's attendance (100%)	Percentage of observance of principles
1	57.5	96.75
2	52.78	95.65
3	27.5	92.5

principles (one hundredth) and do not adhere to the radio (92.5 ± 8.12), of 11 questions, which contain 27.5% of the total questions, each fully adhere to the twelve principles (100%).

4. DISCUSSION

In this research, in the theoretical lessons of radiology 1, 2 and 3, 782 students responded to 611 multiple choice questions, the discrimination, difficulty and correlation coefficient and Millman principles, along with the analyzing the options of the questions, were evaluated. In its interpretation, if D≥ 0.5, the discrimination coefficient of the question is very good, if D≥0.3 and D≤ 0.2 it is moderate, and if D ≤0.19, it is moderate to weak [4].

The discrimination index for radiology 1 is moderate, and for radiology 2 is weak to bad, and for radiology 3 has been strong. It means that questions of radiology 1 have an acceptable extent to be able to distinguish students in the strong group from the weak group. The questions of radiology 2 have had more weakness, therefore 8.3% of the questions with the discrimination coefficient of zero means that they have never been able to differentiate between students of strong and weak groups, and 2.7% with the negative coefficient means that stronger groups are worse off than the weaker group. This kind of questions has underlying disadvantages that should be eliminated or radically revised. This reason is associated with the students in the strong group who have not completely understood or have learned wrongly.

Table 7. Results of questionnaire survey in frequency

Radio	Stem transparency %	Negative specify stem %	Option specified %	Opposite option %	Positive on the stem %	Stalk writing structure words %	Duplicate option %	Stem positive and option %	Vertical options spelling %	Stem positive and option %	Use the phrase all the items in the option %	Use the phrase in the option %
1	100	100	100	97.5	100	90	100	100	95	90	100	100
2	100	100	91.67	97.22	86.11	100	100	100	100	86.11	100	100
3	100	85	95	80	85	97.5	92.5	100	100	85	100	100

According to this common classification, the discrimination index of questions in Radio 1 has a better situation in comparison with its Radio 2.Since the average discrimination index is clear, there is a practical similarity to radio 2. The discrimination coefficient in radio questions 3 is different from radio 1 and 2. Also, 40% of questions in radiology 3 had a strong coefficient, and it is best to distinguish students in the stronger group from the weak group.

The difficulty that is represented by P is the percentage of people who have chosen the correct answer and have a value between 0 and 1. Regarding its interpretation should be said that 50% › $P_{,}$ a hard question, and %75≥ P ≥ %50, a moderate question, and %75 ‹ P is counted as a simple question [4]. As a result, due to the difficulty of radio 1, 2 and 3, all questions with moderate difficulty were above 50%. Also, at least half of the strong and weak groups which are selected can answer correctly. In Radio Lesson 1, only 10% of questions have a difficulty factor above 95% while on radiology 2, %22 have a difficulty above 95% and in Radio 3,2% questions were over 95% which must be revised in these questions.

In correlation analysis, the total score with the answer to each question, which is the coefficient of agreement on the question with the total test. The greater of this correlation coefficient means that the question is more consistent with the set of other questions leading to considering more differences between the strong and the weak. If this coincidence is lower, consistency will be less [4]. In the correlation analysis of the total score with the answer to each question in Radio 1, only 50% of the questions, radiology 2, minimum 33.3% and Radio 3, maximum only 57.5% have been correlated with the total test. Incoherent questions must be rewritten or get rid of them.

Questions with a strong correlation coefficient have a higher value, that is, the answer to this question depends on the strength of the students with their score. It means that stronger students have answered this question correctly and weak students have given it an incorrect answer.

Based on this result that the weakest correlation coefficient is related to in questions of Radiology 2 (33.3%), then these questions should be rewritten. The strongest correlation coefficient is associated with questions of radiology 3 that should be used in other tests. In the evaluation of all Millman principles, the obtained values are acceptable, which is the highest in Radio 1 (96.75) and in radiology 3 is less than the rest (92.5%). 57% of questions of radiology 1 comply fully with the Millman Principles (highest) and radiology 2 and 3 have followed 52%, 72.5%, of the Principles respectively.

In comparison to each of the Millman Indicators, radiology 1 stem transparency, negative stem option, selected option, positive words in the stem, repeating options, the spelling of stem and options, use of expressions of all cases, and none of them are observed 100%. On radiology 2, stem transparency, negative stem option, stem writing structure, repeating options, the spelling of stem and options, showing options vertically, use of the expressions of all cases, and none of them is considered 100%, and on radiology 3, stem transparency, the spelling of stem and options, showing options vertically, using all options and none are followed 100%.

Millman's principles have been observed in questions of Radio 1, 2 and 3. Overall, the findings of this study show that Radio 2 questions have been slightly weaker in the aspect of the above-mentioned indicators and should be considered more clearly than designing other questions.

Shahi et al. did not find any significant difference in the percentages of questions with the proper differentiation between the encyclopedia examinations of the years 2017 and 2018. The results of this study in terms of the percentage of observance of the structural principles of the questionnaire based on the Millman checklist are similar to our research on Radio 1 (96.75), on Radio 2 (95.66%), and Radio 3 (92.5%) at medium level and the clean coefficient of this study was similar to our research, which did not show a significant difference among the discrimination coefficient of the three radio groups.

Zare et al. showed that the level of difficulty and the power of discrimination in all the courses are considerable. Moreover, there was no relationship between the difficulty and the discrimination coefficient. Although the difficulty coefficient for all exams was acceptable and appropriate, the results of the discrimination index showed that for all questions, the quality is weak to very weak. Consequently, it is better to remove these questions from the bank or they should be revised basically. The acceptable and proper level of difficulty in this study was similar

to our research. The results of the poor discrimination index of these questions were similar to those of the radio [13] discrimination factor = 0.23 (versatile 2), the coefficient of discrimination = 0.81 (which was also weak). However, they are not comparable to radio questions in comparison with radio questions [3], the coefficient of discrimination = 0.40.

Meyari et al. [14] found that out of a total of 1239 questions in the study, 63.1% and 76.3% of the questions related to the years of 2008 and 2009 were designed without any form, which was statistically significant. The results of this study are based on the fact that the percentage of observance of Millman structural principles in the average of 63.1% and 76.3% is similar to our research on radiology 1 (coefficient difficulty= 54.97%, Radio 2 (coefficient difficulty = 86.8% 73) and radiology 3, (coefficient difficulty) was 65.73%. In terms of structural errors, 46% of the questions were without a problem and the rest were one or more structural forms.

Regarding structural errors, Haghshenas et al. [7] realized that 46% of questions were without any error and the rest of them had one or more structural errors. The taxonomy and structural errors of the lessons of apprenticeship and physiopathology were significantly less than those of basic science.

In this research, the questions without structural errors were 46%, which were similar to our research on radiology 1 (57.5%), radiology 2 (52.7%) and radiology 3 (73.5%) of the questions without any structural errors. Maybe the professors of Mazandaran University of Medical Sciences have completed courses related to designing the standard questions. Pure Mirza Kalhori found that the average of discrimination index, the test validity, the percentage of questions with minimum structural loss and the closure of tetanus (0.54) were detected in urological questions and the highest coefficient of difficulty (0.67) and the highest increase in the coefficient of differentiation is in psychiatric questions. Due to the increase in the percentage of the coefficient of cleanliness and the coefficient of difficulty in this research, the weakness of the discrimination coefficient in radio theoretical 1 (0.23) and radio 2 (0.81 and radio 3 (40.0) seems to be due to the low quality of the questionnaire in the Medical science field and the education-related to designing multiple-choice questions is essential. Shahi et al., in their study showed that the organization of these courses has been able to increase the quality of the questions to 64%.

5. SUGGESTIONS

The biggest advantage of this project is to provide feedback to the faculty members of Oral radiology department who can take reformative action through the analysis on each of the questions.

Considering the importance of proper assessment of students in medical sciences that is related to the health of the community, it is suggested that these questions be evaluated in other courses to ultimately make a proper assessment of the students. Since the student assessment is one of the components of the evaluation in educational programs, it is suggested that these evaluations be repeated frequently to take into account the educational needs of formative evaluations.

6. CONCLUSION

The quality of the questions based on the principles of Millman in all three radio courses (1, 2 and 3) is standard. It seems that the coefficient of differentiation and the coefficient of difficulty of radio questions is low, especially in Radio 2, which is weak.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Kaveh Tabatabae MS, Behrani Toosi MH, Derakhshan A, Gholami H. Analytical assessment of multiple-choice test at medical school. Journal of medical Education. 2003;2(2):97-91.
2. Jordan. 2004;426-8.
3. Shakoornia A, Khosravi A, Shariati A, Zarei A, Survey on the multiple-choice question of faculty members of the Jondishapor medical university of Ahwaz. The 8th

Nationa Congress of Medical Education. Kerman: Kerman University of medical sciences. 2007;44.
4. Seif A. Educational measurement, assessment and evaluation. 4th Ed Tehran; 2012.
5. The 8th National Congress of Medical Education. Kerman: Kerman University of Medical Sciences. 2007;68.
6. Gholami Vafamehr V, Dadgostarnia M. (Assessment of Effect of proficiency on MCQ in the reliability of clinical medicine query in Esfahan Medical University; 2005.
7. Newell FW, Leopold IH, Adler FH. The multiple-choice question test of the American Academy of Ophthalmology. The American Ophthal. Society Publication. 1970;68:163-170.
8. Mehrens WA, Lehmann IJ. Measurement and evaluation in education and psychology. 4th Ed. New York, Wadsworth Publishing; 1991.
9. Haghshenas M, Vahidshahi K, Mahmoudi M, Shahbazi Nejad L, Parvin Nejad N, Emadi A. (Evaluation of multiple-choice questions in the school of medicine, Mazandaran University of Medical Sciences. The First Semester of 2007. Strides in the Development of Medical Education. 2008;5(2):120-7.
10. Vahid Shahi Kourosh, Mohagheghi Mohammad Ali, Shakeri Stegideh, Sotbouri Mustaheh, Mohtardi Composer. Comparison of some quality indicators of multi-choice questions of the written test of the encyclopedia 6813 and 97. Quality Improvement Unit Testing Specialty Secretariat Secretariat; 2014.
11. Rasolinejad SA, Vakihi Z, Fashion E, Mosayebi Z, monorail R, comparative survey of taxonomies of residents promotion, examination Kashan Medical University; 2006.
12. Zare Shahram, Solati Mehrdad, Hosseini Teshini Saeed. Quality analysis of quiz questions. Conclusion. Hormozgan Medical Journal. Peace 8913. 14(3).
13. Meyari Azam, Biglokhani Mehdi, Zandi Mohammad, Mir Esmaeili Amir Farhang. The effect of educational intervention on the optimization of multi-choice questions design in dental promotion promotion tests. Iranian Journal of Medical Education; 1913.
14. Pourmirza Kalhori R, Rezai M. Roshanpour F (Qualitative analysis of medical promotion exams at Kermanshah University of Medical Sciences. Special Supplement for a 10th National Congress of Medical Education. 2009;335-336.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/54264