Published online 2021 May 17.

Research Article

Analysis of the Orthodontic Exams in Department of Orthodontics of Shahid Beheshti Dental School

Kazem Dalaei¹, Ali Khorsand Nobahar², Mohammad Behnaz³, Mahtab Nouri³ and Fahimeh Anbari^{4,*}

¹Dental Research Center, Research Institute of Dental Research, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

²Dentist, Tehran, Iran ³Department of Orthodontics, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁴Department of Oral Medicine, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

corresponding author: Department of Oral Medicine, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: fahimeh.anbari@gmail.com

Received 2020 November 24; Revised 2021 March 26; Accepted 2021 April 24.

Abstract

Background: Exams are an inseparable part of education, and they are the main tool to evaluate the results of educational process. If based on scientific rules, exams can show the level of achievement to educational goals.

Objectives: This study aimed to evaluate the quality of orthodontic exams in Shahid Beheshti Dental School, Iran.

Methods: In this cross-sectional study, the quality of all written and multiple-choice questions (MCQ) exams of orthodontic department in 2014 - 2015 were evaluated in terms of content validity, structural rules, discrimination index, difficulty index, and taxonomy. To collect the data, the Millman checklist, content analysis table, and standard formula were used. SPSS software version 21 was used to analyze the data. *T*-test, chi square, and independent sample test were used for statistical analysis.

Results: Our results showed that 86.55% of exams were in MCQ form and 13.5% in written form. Content validity was considered in 94.8% of exams. Standard structural rules were observed in 93.61% of questions. Discriminative index and difficulty index were estimated to be 41.6 and 68.8%, respectively. Also, 43.7% of questions were categorized as taxonomy I, 44.38% as taxonomy II, and 11.58% as taxonomy III.

Conclusions: Exams held in Shahid Beheshti Dental School in the 2014 - 2015 academic year were acceptable considering their content validity, structural rules, and difficulty index. However, their discrimination index was low, and they were in poor conditions in terms of taxonomy.

Keywords: Analysis, Orthodontics, Educational Measurement, Shahid Beheshti Dental School, Examination Questions

1. Background

Evaluation is considered as a systematic process of data collection, analysis, and interpretation in order to judge and determine whether the desired objective is achieved or will be achieved, and the extent to which it is achieved (1). If an exam has no desired design in the field of taxonomic selection of questions, observance of structural rules, content validity, and other indices of exams standardization, not only the main role of the exam, which is a complementary and vital component of training, is lost, but also it has negative impacts on learners and devalues the efforts of professors and the educational system; thus, it is necessary to investigate the quality of questions and the standardization of the exam (2).

Accurate understanding of the structure of the exams and the existing problems is necessary to improve the quality of existing programs according to the standards determined by the Education Development Center (EDC) of the university. The results of some studies have shown that professors do not have sufficient skills in designing course exams. As a result, they often use simple and superficial concepts of the course to measure the academic achievement of students, and in most cases, the questions are vague (3).

Several studies have investigated multiple choice questions (MCQ) in universities and higher education institutions worldwide (4-6). The results of these studies showed that some inappropriate questions are used in MCQ exams, and there are serious problems, especially structural problems, in implementation and development of exams. In addition, many instructors teach without sufficient training in measurement and development of exams.

To solve this problem, first, the existing shortcomings and weaknesses are identified through considering the priorities. Then, by holding training courses to introduce the available resources and conducting studies in this

Copyright © 2021, Journal of Medical Education. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/) which permits copy and redistribute the material just in noncommercial usages, provided the original work is properly cited. field, the necessary skills are provided to the teachers.

Considering the importance of evaluation in the higher education system and the need for preparation and empowerment of teachers in preparing and designing appropriate exams, this study aimed to investigate fourchoice theoretical exam questions at the end of the students' rotations in the orthodontics department in the second semester for the academic year 2014 - 2015 to find the possible shortcomings.

In measurement, the attributes or properties of objects and individuals are determined, and the value of those attributes or properties is reported as a number with a digit, therefore it can be said that "measurement is the process that determines that a person or an object has a certain amount of an attribute or properties" (7).

Nitko (2006) defined educational measurement as follows: "how to assign a number (commonly called a score) to an attribute pack with a specific attribute of a person so that the number indicates the degree of that attribute with the attribute that the person has" (8).

Measurement is a process. This process requires a tool called a measuring device. Different tools are used to measure the different attributes of objects and individuals (9).

By definition, then, an exam is a systematic means or method of measuring an example of behavior (10). The more examples of selected and included behaviors in the exam are representative of the desired psychological attributes, the more accurate and reliable the measurement sources are.

An MCQ exam consists of a number of questions, each of which consists of a main part and a number of options (answers), and the exam taker chooses the correct option (answer to the question) from the proposed options (9). These exams can measure most learning outcomes from knowledge to understanding, judging, problem solving, providing practical suggestions, and predicting things. Almost any understanding with the ability measured by other exams - short answer, complete, correct-incorrect, matching, and written - can be measured by MCQ exams (11).

In written exams (essay exams), on the other hand, the examinees present their knowledge in an essay format. Therefore, such exams are most often used in small-scale settings such as educational courses where the number of examinees is limited (12).

The purpose of analyzing exam questions is to investigate each question and determine their accuracy and inadequacies. For analyzing exam questions, the strengths and weaknesses of an exam and the quality of all its questions are determined. Therefore, it is necessary for professors to analyze the questions after each exam and use the results to revise the exam and improve the quality of questions for later use.

The information needed to analyze the questions of an exam are the answers given by the exam takers to each question; so it should be determined how many individuals have chosen the right option in each question, each of the deviation options attracts how many individuals and how many individuals have left them unanswered (8, 13).

The total percentage of subjects who answer a question correctly gives the difficulty index of that question. The larger the difficulty index of a question, the easier that question is, and the smaller the index, the more difficult the question. In general, when the difficulty index is between 0.3 and 0.7, maximum information is provided about the differences among the subjects (14).

Unlike difficulty index, which indicates how easy or difficult a question is for the exam group, the coefficient of discrimination, denoted by d, indicates the strength of the question in distinguishing between the strong group and the weak group of exam subjects, i.e., it shows how much the question can separate the strong group from the weak one.

Each type of measuring tool should have features so that the tool is useful for the purpose for which it was made. The best of these features are validity and reliability. Validity is a term that refers to the objective that the exam is designed for. It is a valid exam that is suitable for measuring what is intended. Content validity indicates the extent to which the sample questions used in an exam represent a comprehensive set of possible questions that can be prepared from the content with the subject matter. In order to prepare a good academic achievement test, the exam questions should be a complete sample of the objectives and content of the course (15).

The reliability of a measuring device refers to the accuracy of that device. An exam is reliable if we give it to a single group of individuals several times in a short period of time, and the results should be close to each other (16).

2. Methods

This descriptive cross-sectional study was conducted on all MCQ and written exams of theoretical orthodontics courses 1, 2, and 3 at the end of the second semester at the Orthodontics Department of Shahid Beheshti Dental School in the academic years of 2014 - 2015.

The information was collected through checklists, calculations, expert consensus, interviews, and group discussions.

The coefficient of discrimination and difficulty index of the questions were calculated by the following formula. Coefficient of discrimination = (Number of correct choices of the higher group- Number of correct choices of the lower group)/Total number of individuals

Difficulty index = Number of correct choices / Total number of individuals

The list of students' scores was sorted from higher to lower. One-third of the students who scored higher than the range were in the higher group, and one-third of the students who scored lower than the range were in the lower group.

The list of scores of each exam was provided to the researcher through training, which was copied from the list of original scores. The names of the students were removed, and only scores were available in the copied lists. For each exam, the sheets of the higher and lower groups were separated, and for each question, the number of correct choices in each group was counted separately and placed in the above formulas to obtain the coefficient of discrimination or difficulty index for each question. It is easy to measure the right choices for multiple choice or true and false questions.

The degree of observance of structural rules in questions and exams (Table 1) (17) was measured in two separate tables.

The cognitive level (taxonomy) of the questions and the content validity of the exams were measured by three faculty members of the orthodontics department.

Content validity determines that the test questions are related to the educational objectives of the course and it is 100% when all the questions are related to the educational objectives and there is no question that has not considered a goal or is not related to the objectives or content of the course plan (15).

The content analysis table (Table 2) was provided to three faculty members of the orthodontics department. In this table, four-choice questions are placed along with the options of each question. Two extra columns were provided to determine the content coverage and the cognitive level of each question.

The data obtained from the study were analyzed using SPSS software version 21 and *t*-test, chi square, and independent sample test. The tables of frequency and frequency percentage and bar graphs were used to review the study results.

3. Results

In this study, a total of 147 questions and the sheets of 249 students were investigated. Table 3 shows the frequency and distribution of students and questions.

The overall observance of structural rules was 93.6%, which was 91.2% in 2014 and 96% in 2015 (P = 0.8).

The content validity of exams considering the lesson objectives was 94.8% totally, which was 96.6% in 2014 and 93% in 2015 (P = 0.7).

The difficulty index was 68.4% in general, that was 69.5% in 2014 and 67.3% in 2015 (P = 0.5).

The overall discrimination index was 0.41, which was 0.42 for exams held in 2014 and 0.4 in 2015 (P = 0.8).

The distribution and percentage of all variables are shown in Table 4.

4. Discussion

Our results showed that most of the orthodontic exams are held in the form of MCQ (86.55%), and only a small part (13.5%) is in the form of written questions. Interestingly, this distribution does not exist in all exams, and in some of the exams, only MCQs are used, which indicates that most professors use objective questions to measure students' academic achievement and use subjective and written exams less. The objective exams are those in which both the question and the answer are given to the exam takers, and the exam taker decides on the given answers. These exams usually measure the ability to recognize, while written exams generally measure the ability to recall. Although these two processes of recall and recognition are psychologically closely related and constitute aspects of memory, they differ in terms of practical and applicable abilities. Therefore, it can be said that most orthodontic exams measure students' knowledge to the extent of recognition. However, it is better to use written and objective exams (multiple-choice questions) at the same time to better identify what students have learned.

According to the results of data analysis, the average rate of observing structural rules in orthodontic exams was 93.61%, which is not significantly different from the average for exams in all departments in 2000-2001 by Baharvand et al. (18) (96.2%); in general, this shows the high quality of orthodontic exams in terms of structure. The case that was not observed in terms of structural validity more than others was related to the use of negative words and verbs, as well as using the option of 'all or none of the options' in the answers.

The content validity, expressed here as content relevance, was evaluated against the course plan which was planned by the department.

The content validity of the studied exams was 94.8%, which shows that unrelated questions related to the taught chapters are rarely included in the exams, and the educational objectives are well covered. This result is also consistent with Baharvand's (18) as well as Kazemi's (19) studies. They evaluated 1013 questions gathering from

able 1. Investigating the Exam According to Miniman Principles			
Type of Question/Index	% of Observance	% of Non-observance	Total Cases
Written			
Avoid using the words of who and when			
True/false			
The text length of the correct and incorrect questions is the same			
Multiple choice			
Match the length of the options			
Short options			
Definite and prominent negative words			
Do not use conflicting options, one of which is correct			
Failure to design double negative questions			
Just one correct answer			
Deviation option in terms of length, structure, and words used			
Do not use the same words in the questions and options			
Do not use A and B or B and C			
Do not repeat a phrase in all options			
Do not use all or none of options			
All questions			
Do not use absolute adverbs like only, never, and always and words like often and sometimes			
Use correct grammatical rules			
Do not use spaces at the beginning or middle of the question text			
Do not use abbreviations			
No writing errors			
No misspellings			
The main content is completely in the main text of the question			
A clear question			
Evaluate a specific goal			
Not irrelevant long text of the question			
Using positive verbs			

Question	Validity	Taxonomy	According to the Lesson Objectives
	Difficulty index	Taxonomy I	
The text of the questions is not displayed in this table due to the fidelity and safety of the questions	Coefficient of discrimination	Taxonomy II	
	Structural validity	Taxonomy III	-

18 exams, and their results showed a content validity of 92.38%.

Validity refers to the objective that the exam is designed to achieve. A valid exam is the one that is suitable for measuring what is desired.

Many sets of questions can be prepared in a subject. Content validity indicates to which extent the sample of questions used in an exam represents this comprehensive set. In order to make the exams valid, the exam questions should be a complete example of the objectives and content of the course. One of the factors that affect the validity of the exam is the quality of the exam questions. A question that is prepared according to the rules in the field of questioning adds to the validity of the exam (15).

The Name of the Evam	Educational Voar	Total Number of	Number o	Number of Students		
The Name of the Exam	Educational Ical	Questions	MCQ	Written	Number of Students	
Orthodontics 1	2014	20	20	0	45	
Orthodontics 2	2014	20	20	0	44	
Orthodontics 3	2014	34	30	4	26	
Orthodontics 1	2015	23	23	0	53	
Orthodontics 2	2015	40	34	6	36	
Orthodontics 3	hodontics 3 2015		14	12	45	

Table 3. Frequency and Distribution of Students and Questions

Table 4. Details of All Variables According to Exams (in Percentage)

The Name of the Exam	Educational Year	Content Validity Index	Structural Validity Index	Difficulty Index	Discrimination Index	Number of Questions		Тахопоту		
						MCQ	Written	I	II	Ш
Orthodontics 1	2014	95	87.85	65.7	0.48	20	0	55	25	20
Orthodontics 2	2014	95	91.7	67	0.44	20	0	55	40	5
Orthodontics 3	2014	100	94	75.6	0.35	30	4	36.7	42.9	21.4
Orthodontics 1	2015	91.3	94.94	74.1	0.35	23	0	47.8	52.2	0
Orthodontics 2	2015	91.7	96.2	63.3	0.42	34	6	32.4	52.9	14.7
Orthodontics 3	2015	95.8	97	64.6	0.43	14	12	35.7	42.9	21.4

The average difficulty index of the studied exams in this study was 68.38%. The orthodontic exam of the second semester of 2014 - 2015 had the highest difficulty index, which can be considered due to the small number of MCQ exams, that were the purpose of this study. Also, orthodontic exam 3 had the lowest difficulty index.

It can be said that the difficulty index of the exams was appropriate, and the exams were not very difficult or very easy. These results are consistent with the results of Abbasi et al. (20) in which only 27.6% of questions had an appropriate range of difficulty coefficient. However, they are inconsistent with Imam Jome et al. (21) results in which 62% of the questions were considered as easy.

One of our expectations of standard exams is that the exam takers got scores which are distributed over a range. The wider the distribution, the better is the exam. In other words, the higher the variance of the scores obtained from the standard exam, the better the exam. Considering the difficulty index obtained from this study, it can be said that almost all the studied exams were appropriate in terms of the distribution of scores, and they were suitable and standard in this respect (variances).

A larger coefficient of discrimination indicates the distinctive power of the question, and the closer this number is to one, the more appropriate it is. It can be said that the coefficient of discrimination of the studied exams was low (with an average of 0.4116), which shows that the studied questions do not have the power to distinguish between strong and weak students.

The students of Shahid Beheshti Dental School are considered the best ones and also at the same level of knowledge in the entrance exam. This can be assumed the reason for the low coefficient of discrimination, but the distribution of students' scores in the exams and reports obtained from teachers rejected this claim. So, it can be concluded that in terms of the distinctive power, question design is not appropriate. These results are consistent with the results of Abbasi et al. (20) in which 87.6% of the questions had the coefficient of discrimination of less than 0.5. Also, in a study by Baharvand et al. (18) conducted on the written exams of theoretical courses, the coefficient of discrimination was 22.3% in all departments, which shows that the orthodontic department exams have much higher distinctive power than the theoretical exams of other departments.

The percentage of taxonomy of the questions was nonuniformly distributed. Approximately 90% of the questions were in the Taxonomy I and II. In the orthodontic exam of the second half of 2013 - 2014, no question was placed in Taxonomy III group. Also, no significant relationship was observed between the high percentage of questions and the cognitive level of Taxonomy III, and the difficulty index of the exam. This might be due to the study population of our study. We used only MCQ exams for calculating the difficulty index, while most questions in the level of Taxonomy III were in the form of written question.

In general, the results show that most of the questions designed in orthodontic exams of Shahid Beheshti Dental School measure most of the students' mental archives, and the number of questions for which the student needs to be analyzed is small. This finding is consistent with the study by Shakournia (3), in which 90% of MCQ exams were in Taxonomy I and II. It is also consistent with the study by Baharvand et al. (18), in which the percentage of questions with this taxonomy was 84.6%. Also, with increasing students' educational level and the expansion of the taught materials, the possibility for professors to design questions with higher cognitive levels (taxonomy III) is increased, and more questions were observed at this level in the theoretical orthodontic exam 3 in both semesters of 2013 - 2014 and 2014 - 2017. Taxonomy I and II generally investigate memory and recall, and Taxonomy III analyzes and combines the information of the exam subject with the practical power of the information. From Anderson/Krathwohl's point of view (2001), Taxonomy I measures knowledge that is less involved in understanding at this level, and the only memorization is investigated. At the cognitive level, Taxonomy II requires a greater understanding of superficial awareness of information in memory. Understanding the content causes them not to be forgotten. At a higher cognitive level, Taxonomy III, the lessons learned are more in-depth than the previous two ones, and we need to combine previous information and lessons learned.

Also, with comparing the index between similar exams of 2013 - 2014 and 2014 - 2017, the only index that had a significant difference was the structural validity index in the theoretical orthodontic exam (P = 0.01) and theoretical orthodontics 2 (P = 0.06). In both cases, the structural validity was increased in 2014 - 2015 exams.

4.1. Conclusion

According to our results, MCQ/written exams held in the orthodontics department of Shahid Beheshti School of Dentistry in academic years of 2013 - 2014 and 2014 - 2015 were at a relatively good level in terms of observing indices for measuring the quality of exams. While the content validity, observance of structural rules, and difficulty index of the exams were appropriate and acceptable, the discrimination coefficient and cognitive level of the questions (taxonomy) were at a low level, which requires more attention.

Footnotes

Authors' Contribution: Study concept and design, K.D.; Acquisition of data, K.D.; Analysis and interpretation of

data, A.K.; Drafting of the manuscript, A.K.; Critical revision of the manuscript for important intellectual content, M.N.; Statistical analysis, M.N.; Administrative, technical, and material support, M.B.; Study supervision, F.A.

Conflict of Interests: The authors declare that there is no conflict of interests.

Funding/Support: Not applicable.

References

- 1. Bazargan A. [Measurement tools for educational evaluation]. [Educational assessment]. 5th ed. Tehran, Iran: SAMT; 2001. p. 79–91. Persian.
- 2. Seif A. [Educatianal measurement, assessment, and evaluation]. 4th ed. Tehran, Iran: Agah; 2007. p. 60–132. Persian.
- 3. Shakoornia A, Khosravi A, Shariati A, Zarei A. [Survey on multiple choice question of faculty members of Jondi Shapour Medical University of Ahvaz]. *The 8th National Congress of Medical Education*. Kerman, Iran. Kerman University of medical sciences; 2009. p. 44–59. Persian.
- Tarrant M, Knierim A, Hayes SK, Ware J. The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Educ Today*. 2006;26(8):662–71. doi: 10.1016/ji.nedt.2006.07.006. [PubMed: 17014932].
- 5. Sepasi H, Atari YA. [The study of psychometric characteristics of Shahid Chamran University faculty members' final test scores]. *J Educ Psychol*. 2006;**12**(4):1–20. Persian.
- Haghshenas M, Vahidshahi K, Mahmudi M, Shahbaznejad L, Parvinnejad N, Emadi A. [Evaluation of multiple choice questions in the school of Medicine Mazandaran University of Medical Sciences the first Semester of 2007]. Stride Dev Med Educ. 2009;5(2):120–7. Persian.
- 7. Gay LR. Educational evaluation and measurement. 2nd ed. USA: Macmillan Library Reference; 1991. p. 13–8.
- 8. Nitko JA. Educational assessment and evaluation. 3rd ed. New York, USA: Pearson College Div; 2001. p. 5–26.
- 9. Seif A. [Measurement, educational assessment and evaluation]. 3rd ed. Tehran, Iran: Doran; 2007. Persian.
- Grounland NE, Linn RL. Measurement and evaluation in teaching. 1st ed. USA: Pearson College Div; 1990. p. 3–7.
- 11. Ebel RL. Essentials of Educational Measurement. 3rd ed. New Jersey, USA: Prentice-Hall; 1979.
- Nnodim JO. Multiple-choice testing in anatomy. *Med Educ*. 1992;**26**(4):301–9. doi: 10.1111/j.1365-2923.1992.tb00173.x. [PubMed: 1630332].
- 13. Whitney DR, Sabers DL. Improving essay examinations: Use of item analysis. 4th ed. New York, USA: Wadsworth; 1970. p. 1–3.
- Glover JA, Bruning RH. Educational psychology: Principles and applications. Virginia, USA: Scott, Foresman/Little Brown Higher Education; 1990.
- 15. McBurney D, Middleton P. *Research methods*. 1st ed. California, USA: Brooks Cole Publishing company; 1998. p. 1–3.
- 16. Bryman A. Quantitive data analysis. 3rd ed. London, UK: Routledje; 1999. p. 23-5.
- 17. Orangi AM, Dorani K. [Developing a social studies achievement test for high school students based on item-response theory (IRT)]. Journal of Psychological Models and Methods. 2010;1(1):1–13. Persian.
- Baharvand M, Hoseinzadeh M, Jaberiansari Z, Abbaszadeh E, Mortazavi H. [Evaluation of 30 structural validity and content quality indices of theoretical dental exams]. J Mash Dent Sch. 2015;38(4):291–302. Persian.
- Kazemi A, Ehsanpour S. [Item analysis of core theoretical courses exams for midwifery students in Isfahan University of Medical Sciences]. *Iran J Med Educ.* 2011;10(5 (29)):643–50. Persian.

- Abbasi A, Kamkar A. [Evaluation of the end-of-semester exams of different courses of the School of Nursing and Midwifery in the year 1984-85]. Research And Science Journal of Yazd University of Medical Sciences. 2007;1(1):13–9. Persian.
- Imam jome M, Zahedifar F. [Analysis of multiple choice questions exams in health disciplines of Qazvin University of Medical Sciences]. *Scientific Journal of Qazvin University of Medical Sciences*. 2011;15(4):110– 8. Persian.