

AJDR Avicenna Journal of Dental Research

Avicenna J Dent Res, 2024; 16(1):9-15. doi:10.34172/ajdr.1595

http://ajdr.umsha.ac.ir



Original Article



Development of an Instrument for the Online Evaluation of Orthodontic-Treated Cases: A Pilot Study

Amirfarhang Miresmaeili¹, Hana Salehisaheb²

¹Department of Orthodontics, School of Dentistry, Hamadan University of Medical Sciences, Hamadan, Iran ²Department of Orthodontics, School of Dentistry, Kurdistan University of Medical Sciences, Kurdistan, Iran

Article history: Received: February 26, 2022 Accepted: May 2, 2022 ePublished: March 29, 2024

*Corresponding author: Hana Salehisaheb, Email: hanasalehisaheb@gmail.com

Abstract

Background: It is necessary to create a sound method for the optimal evaluation of treated cases of orthodontic board exam candidates. This cross-sectional study aimed to design and validate a new method for the national orthodontic board exam using a web-based program.

Methods: Complete documents of 10 patients randomly selected from a pool of the previously presented cases at the national board examination were entered into a web-based program called "the Orthoboard". The documents were arranged according to the European *Board of Orthodontist* Standards, and 15 related questions were asked based on the *index* of complexity, outcome, and need index and the American Board of Orthodontics evaluation standards. A customized grading system was used for the finalized questionnaire. They were asked to be evaluated by 10 orthodontists (5 with less and 5 with more than 10 years of experience). The content validity of the questionnaire was analyzed by the content validity index and content validity ratio. The reliability of this questionnaire was measured using the Kappa statistical test. **Results:** Most evaluators do not consider it necessary to ask questions about the diagnostic resume, problem list, treatment plan, and treatment resume. Crowding after treatment and crossbiting before and after treatment had the best interexaminer reliability. The least agreement among examiners was between the pre-treatment aesthetic index and the pre-treatment buccal segment relationship.

Conclusion: Inter-examiner reliability was lower than expected, indicating that the orthodontic board test scoring is too subjective. The addition of a 3-dimensional cast is recommended for better objective evaluation.

Keywords: Specialty boards, Web-based software, Software validation, Reproducibility of results, Records



Please cite this article as follows: Miresmaeili A, Salehisaheb H. Development of an instrument for the online evaluation of orthodontic-treated cases: a pilot study. Avicenna J Dent Res. 2024; 16(1):9-15. doi:10.34172/ajdr.1595

Background

Board certification has been widely accepted as a means of improving the quality of medical and dental care (1). There are many reasons in support of achieving board certification, one of the most important of which is self-satisfaction. The importance of certification in the eyes of the general public has also increased (2). This test in Iran includes written, oral, and practical parts. Although the comprehensive software for designing test questions (Najma) was used in the written board exam in the 31st exam of the specialized dental encyclopedia in September 1994, the participants' uncertainty and lack of knowledge of the content of the oral exam and the expected capability and level of individual skill in diagnosis and treatment plan, as well as the clinical information of the participants, are among the weaknesses of the board exams in dentistry (3).

The process of introducing new digital technology

innovations into teaching, learning, and assessment has been studied for many years (4). Dental educators have sought for the last decade to integrate computers into the dental curriculum (5). New assessment tools (e-assessments) must be designed because of the creation of new e-teaching and e-learning systems (6). The development of these electronic assessments benefits both students and their assessors because it "improves the reporting, storage, and transmission of data related to public and internal assessments" (7). Especially during the coronavirus disease 19 epidemic, using electronic innovations as much as possible can be an effective step toward improving education and evaluation.

Various methods of orthodontic case evaluation include the American Board of Orthodontics (ABO), European *Board of Orthodontists* (EBO), *index* of complexity, outcome, and need (ICON), and the like. However,



© 2024 The Author(s); Published by Hamadan University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

none of these methods are completely suitable for online evaluation web-based programs. The ICON index is often used to examine patients' dental casts (8,9) and does not include radiographic assessments such as root paralleling. The ABO discrepancy index measures the complexity of cases (10). By combining these indexes, the success rate of treatment can be evaluated while evaluating the complexity of malocclusion.

Computer-based patient records are not new; however, they have evolved significantly in the past ten years because of the emergence of the Internet and related technologies (11,12). Orthoboard is a program designed to simplify and standardize the oral part of a specialized orthodontic board exam on the web. It has specific sections for entering diagnostic information, treatment

plans, radiographic images, and photographs of patients (Figure 1). The arrangement of each patient's information is set according to EBO standards (13). The users of this program are specialized assistants and professors. Access to the program is possible through the *www.orthoboard. ir* web address. Assistants can upload their patients' diagnostic documents into the relevant sections of the program. These documents can be viewed and reviewed by professors. On the page dedicated to professors, there are questions about each of the patients' diagnostic documents based on ICON and ABO indices. There is also a guide on how to answer the questions. Professors give scores to each patient's records by answering these questions (Figure 2).

Reliability is the degree of correlation between the

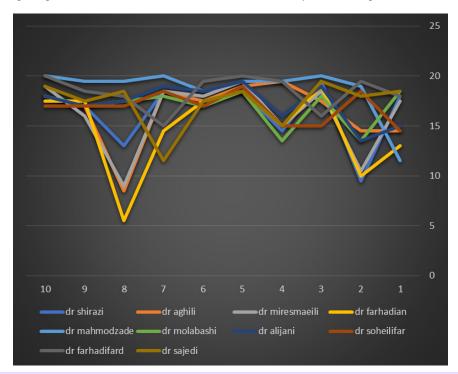


Figure 1. Location of Diagnostic Records in the Left Column of the Orthoboard Program



Figure 2. Questions on the Professors' Dedicated Page to Evaluate Intraoral Photographs

results of two or more measurements that are performed independently and separately. Numerous methods have been proposed regarding quantifying experts' degree of agreement regarding the content relevance of an instrument. These include, for example, averaging experts' ratings of item relevance and using a preestablished criterion of acceptability (14), using coefficient alpha to quantify agreement of item relevance by three or more experts (15), and computing a multi-rater kappa coefficient (16).

The Orthoboard program was designed to create a sound method for the optimal evaluation of treated cases of orthodontic board exam candidates, and the purpose of this study is to evaluate the validity and reliability of this web-based instrument.

Methods

Ten patients who had previously been presented in the national board exam were selected from a pool of 400 documents. It was attempted to select all types of case categories presented in the board exam to identify the shortcomings related to the evaluation of each type of case, if any. The demographic characteristics of the sample are provided in Table 1.

Ten judges (5 with less than 10 years of experience and 5 with more than 10 years of experience) were selected from the orthodontic departments of different faculties. The selection was made based on familiarity with the national board exam and their interest in cooperation. The judges' information is presented in Table 2.

Online instructions for using the Orthoboard program, individual username and password, grading method, and validity questionnaire were emailed to each examiner separately. Then, they were contacted, and the method of using the program and the scoring method were explained once, and possible ambiguities were clarified accordingly.

To evaluate the quality of the treatment, diagnostic information, treatment plan, and radiographic and photographic images of 10 patients who had complete documents before and after treatment were entered into

Table 1. Sample Demographics

Variable					
Gender, No. (%)					
Male	4 (40)				
Female	6 (60)				
Patient age, Mean (SD)	18.4 (5.69)				
Treatment time, Mean (SD)	1.96 (0.3)				
Extraction pattern, No. (%)					
Nonextraction	3 (30)				
Extraction	7 (70)				
Angle class, No. (%)					
Cl II	7 (70)				
Cl III	3 (30)				

Note. SD: Standard deviation.

the Orthoboard program based on EBO standards. The questions asked about each of the patients' diagnostic records were included in the program. These questions are based on the ICON index and evaluation standards in ABO.

ICON index questions included these items as (17):

- 1. Aesthetic evaluation in the frontal occlusal photo
- 2. Evaluation of crowding/spacing in the maxillary occlusal photo
- 3. Evaluation of crossbite in the frontal occlusal photo
- 4. Incisor open bite/overbite evaluation in the frontal occlusal photo
- 5. Evaluation of buccal segments in the left and right occlusal photos

Items of the ABO discrepancy assessment index included:

- Root paralleling in the final orthopantomography (OPG)
- 2. Root resorption in the final OPG
- 3. Mandibular plane angle change in superimposition
- 4. Lips to E-line relationship in the superimposition

Suggested Scoring System

A total of 15 questions with 20 points are included in this scoring system. Due to more objectivity, 16 out of 20 points were assigned from the ICON index, and 4 points were assigned from the ABO questions.

Based on the pre- and post-treatment photographs and according to Table 3, the evaluators assigned scores to the ICON index questions. Then, the score of each question was multiplied by its respective weight. It is 7, 5, 4, and 3 for aesthetic index, crowding/spacing and crossbite, open bite/overbite, and an anterior-posterior buccal segment, respectively. Next, the patient's degree of improvement was calculated based on the following formula:

Pre-treatment score – $(4 \times post-treatment score)$

The grade was in one of these improvement grades, including greatly improved, substantially improved, moderately improved, minimally improved, and not-

Table 2. Demographic Data for Judges

	Age	Gender	Years in Practice	Years in Teaching
Judge 1	74	М	42	42
Judge 2	52	М	25	25
Judge 3	52	М	25	25
Judge 4	52	F	25	25
Judge 5	40	М	12	12
Judge 6	37	М	7	7
Judge 7	33	F	6	6
Judge 8	32	F	5	5
Judge 9	31	F	3	3
Judge 10	34	М	6	4
Average	43/7		15/6	15/4

improved or worse, which were assigned 16, 12, 8, 4, and 0 points, respectively.

Each ABO question was given a score of 1. The answers to these questions were in the forms of "poor", "fair", and "good", which include 0-, 0.-5, and 1-point bars, respectively. The definition of these items is given in Table 4.

Statistical Analysis

The study data were collected and analyzed using Statistical Package for the Social Sciences (SPSS) software, version 24

(SPSS Inc., Chicago, IL, USA). The kappa statistical test was used to evaluate the reliability of the scores assigned to each patient and the reliability of the scores assigned to each question. The content validity of the questions was analyzed by content validity index (CVI) and content validity ratio (CVR) statistical tests.

Results

Figure 3 shows the evaluators' agreement with the studied patients. Most of the professors agreed on cases 5 and

Table 3. ICON Index Characteristics and its Scores

	Score	0	1	2	3	4
Aesthetic	1–10 as judged using IOTN-AC				-	
Upper arch crowding	Score only the highest trait either spacing or crowding	Less than 2 mm	2.1–5 mm	5.1–9 mm	9.1–13 mm	13.1–17 mm
Upper spacing		Up to 2 mm	2.1–5 mm	5.1–9 mm	>9 mm	
Crossbite	Transverse relationship of cusp to cusp or worse	No crossbite	Crossbite present			
Incisor open bite	Score only the highest trait either open bite or overbite	Complete bite	Less than 1 mm	1.1–2 mm	2.1–4 mm	>4 mm
Incisor overbite	Lower incisor coverage	Up to 1/3 tooth	1/2 – 2/3 coverage	2/3 up to fully covered	Fully covered	
Buccal segment anteroposterior	Left and right added together	Cusp to embrasure relationship only, Class I, II, or III	Any cusp relation up to but not including cusp to cusp			

Note. ICON: Index of complexity, outcome, and need; IOTN: Index of orthodontic treatment need; AC: Aesthetic component.

Table 4. Definition of ABO Question Options

	Poor	Fair	Good
Root paralleling	More than 2 roots in contact	2 roots in contact	Overall root paralleling
Root resorption	Greater than 1/4 of root length	Up to 1/4 of root length	No resorption or slight blunting
Lips to E-line	Initial better than the final	Final better than the initial	Ideal lips to E-line relation
MP angle change	Increase in MP angle > 1 degree	No change in the MP angle	Decrease in the MP angle

Note. ABO: American Board of Orthodontics.



6. The agreement on cases 3, 9, and 10 was appropriate, but on cases 1, 2, 4, 7, and 8, there was a significant disagreement. In general, professors with more than 10 years of experience assigned lower scores to the patients, which was especially observed in patients 1, 2, and 8.

The scores assigned by Assessor No. 1 were generally compared with those of other assessors. The reason for choosing evaluator No. 1 as the evaluation standard was his experience. Based on this analysis, it was shown that the agreement of evaluator 1 was highly good compared to No. 3, moderate compared to evaluators 4 and 6, and weak compared to evaluators 2, 5, 7, 8, 9, and 10 (Table 5).

Moreover, the level of agreement of evaluator 1 with other evaluators was examined in terms of each question. Most evaluators agreed on crowding after treatment and then cross-biting after and before treatment. The least agreement was in the pre-treatment aesthetic index, pre-treatment buccal segment, pre-treatment overbite, and lips to E-line relationship. Other questions fell between these two categories (Table 6).

CVR was utilized to assess the "necessity", "clarity", and simplicity" of the questions. CVR values greater than 0.62 were accepted since the number of evaluators was 10. CVI has also been employed to check the "relevance" of the questions, and the content validity of the question was confirmed if the CVI score was higher than 0.79. In general, most evaluators did not consider it necessary to score questions 6 (case resume), 7 (problem list), 8 (treatment plan), and 9 (treatment resume). In addition, 30% of them did not consider it necessary to give a score for question 15 (lips to E-line). The clarity and simplicity of question 5 (buccal segment) were not acceptable to most evaluators (Table 7).

Discussion

In designing this experiment, we attempted to evaluate the

appropriateness and reliability of the electronic scoring of the orthodontic oral board test. There was a favorable agreement between professors with more than 10 years of experience and those with less than 10 years of experience, except for patients No. 1, 2, and 8. The more experienced professors assigned lower scores to these patients. The highest agreement was observed in patients 5 and 6. It seems that the reason that has led to the most controversy was the aesthetic index of occlusion.

The Orthoboard program is designed to facilitate the oral part of the orthodontic board exam. The board exam participants can enter complete patient documents into this program, thus eliminating the need to bring patients' physical documents to the test site. These records can be

Table 7. Content Validity of Questions

Questions	Necessity (CVR)	Relation (CVI)	Clarity (CVR)	Simplicity (CVR)
Q1	100%	1	80%	100%
Q2	100%	1	100%	100%
Q3	100%	1	60%	100%
Q4	100%	1	80%	80%
Q5	100%	0.9	-60%	-40%
Q6	-60%	1	100%	100%
Q7	-60%	0.9	100%	100%
Q8	-40%	0.9	100%	100%
Q9	-60%	1	80%	80%
Q10	60%	1	100%	100%
Q11	60%	1	100%	100%
Q12	60%	1	80%	100%
Q13	100%	1	80%	100%
Q14	80%	1	60%	80%
Q15	40%	0.8	60%	60%
Q16	60%	0.8	60%	80%

Note. CVI: Content validity index; CVR: Content validity ratio.

Table 5. Reliability of Scores Assigned to Patients by Assessor No. 1 Compared to Other Assessors

	Kappa coefficient									
	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7	Rater 8	Rater 9	Rater 10	
Rater 1	0.36	0.8	0.5	-0.1	0.52	0.37	0.16	-0.19	0.26	

Table 6. Reliability of the Scores Assigned to Each Question by Assessor No. 1 Compared to Other Assessors

	Kappa Coefficient													
	Esthetic 1	Crowding 1	X bite 1	Overbite 1	Buccal 1	Esthetic 2	Crowding 2	X bite 2	Overbite 2	Buccal 2	Paralleling	Resorption	Lips to E-line	MP Angle
Rater 2	0.024	0.383	0.615	0.375	0.211	1	1	0.565	0.615	0.42	0.437	0.167	0.028	0.6
Rater 3	0.079	0.268	0.583	0.194	0.25	0.231	1	0.8	0	0.429	0.032	0.412	0.123	0.455
Rater 4	0.259	0.302	0.783	0.516	0.231	0.706	1	1	0.615	0.437	0.565	0.024	0.355	0.8
Rater 5	-0.053	0.474	0.4	-0.061	0.114	0	1	0.615	0	0.394	0.31	0.259	0.167	0.111
Rater 6	0.041	0.419	0.615	0.219	0.146	0.615	1	1	-0.154	0.375	0.206	0.455	0.153	0.636
Rater 7	0.178	0.467	0.783	0.394	0.342	0.615	1	0.615	1	0.254	0.508	0.756	0.091	0.25
Rater 8	0.375	0.5	1	0.394	-0.098	0	1	1	0	0.254	0.18	0.118	0.344	0.455
Rater 9	0.129	0.73	0.444	0.531	-0.098	0	1	1	0	0.677	0.333	0.706	0.136	0.176
Rater 10	0.118	0.213	0.615	0.265	0.342	0.412	1	0.615	0	0.524	0.825	0.63	0.118	0.4

damaged or lost when dispensed to students and need to be stored and maintained every year. Storing records in digital form is an option to overcome problems with handling hard-copy records (5). Further, using the Orthoboard program, the presented records are evaluated more accurately, saving the time required for the oral part of the board test.

According to the study, more experienced professors generally assigned lower scores to the cases presented. It seems logical that more experienced professors use more accuracy and rigor in examining patients. The board professors are generally among the most experienced professors in their field and are probably more similar in terms of the degree of leniency than the conditions used in this study, where both more and less experienced professors were utilized. Therefore, if only board professors were involved in the study, the degree of agreement would probably be higher.

Due to the very high diversity of malocclusions, none of the existing indices, including the standard ICON index, can fully cover this diversity, and this causes the aesthetic part to be determined subjectively. In the ICON standard index, the aesthetic is assigned the highest coefficient, and therefore, even a small difference in the score assigned to this part can lead to a significant difference in the final score. Accordingly, it is probably better to reduce the aesthetic coefficient in the proposed scoring system to achieve greater reliability.

The other criterion that led to differences in scores was crowding/spacing. The reason for the disagreement in this regard was that some professors have considered the extracted tooth space as spacing and some have not. Providing more complete and detailed explanations can reduce the evaluators' disagreement on this question.

In the help table, the score assigned to each overbite/open bite value is specified, but the reason for the disagreement is that some evaluators have considered other anterior teeth to determine overbite in addition to the central. Providing more detailed explanations in the help file can lead to more agreement among the evaluators.

Lateral photographic images used to assess buccal segment occlusion are angled and do not provide a completely direct view of these areas; thus, it is impossible to accurately assess buccal occlusion using these images. Adding 3-dimensional (3D) images of dental models to the Orthoboard program is a great solution to overcome this problem.

A detailed definition is provided for each of the ABO index criteria in the relevant guide table. It seems that the reason for the disagreement is the consideration of personal opinions by the evaluators in scoring these questions. For example, in examining the root paralleling, scoring should be based on the root contacts in panoramic radiographs, but some professors believe that the position of the crown and marginal ridges is more important, and if the position of the crowns is appropriate, they have assigned a complete score to this part, despite the root contacts.

In our proposed scoring system, an improvement grade was used instead of the scores of each criterion directly, allowing evaluators some flexibility in scoring. If there is a slight difference in the scores related to each of the ICON criteria, it may not affect the improvement grade or the final score.

In cases 5 and 6, the scores assigned by the evaluators to the aesthetic index were highly similar. As mentioned earlier, this section has the highest score coefficient in the proposed scoring system. In addition, the initial severe malocclusion has improved significantly, and despite slight differences in scores assigned to other ICON index questions, there has been no difference in the improvement grade, and the final score of these two cases has shown high agreement among evaluators.

Based on the results of the CVR and CVI tests, most evaluators did not consider it necessary to assign a score to the diagnostic resume, problem list, treatment plan, and treatment resume. In orthodontic treatment, there may be several methods to treat a particular patient, and the treatment plan is selected based on the priority given by each orthodontist to the patient's problems. The most important issue is achieving the final desired result. In our proposed scoring system, no score was assigned to these parts.

According to CVR, the simplicity and clarity of the buccal segment question have not been confirmed. As mentioned earlier, this is due to the lack of a perfectly perpendicular view of the buccal segment in lateral photographic images, and this problem can be overcome by adding 3D images of dental models to the Orthoboard program.

Conclusion

In summary, inter-examiner reliability was lower than expected, indicating that orthodontic board test scoring is too subjective. Hence, the addition of 3D models is recommended for better objective evaluation.

Acknowledgments

The authors gratefully acknowledge the support of this work by the Vice-chancellor for Research and Technology, Hamadan University of Medical Sciences, for financial support.

Authors' Contribution

Conceptualization: Amirfarhang Miresmaeili.

Data curation: Hana Salehisaheb.

Formal Analysis: Amirfarhang Miresmaeili, Hana Salehisaheb.

Funding acquisition: Amirfarhang Miresmaeili.

Investigation: Hana Salehisaheb.

Methodology: Amirfarhang Miresmaeili.

Project administration: Amirfarhang Miresmaeili.

Project administration: Amirfarhang Miresmaeili. **Resources:** Amirfarhang Miresmaeili, Hana Salehisaheb. **Software:** Amirfarhang Miresmaeili, Hana Salehisaheb.

Supervision: Amirfarhang Miresmaeili.

Validation: Amirfarhang Miresmaeili, Hana Salehisaheb.

Visualization: Amirfarhang Miresmaeili. **Writing-original draft:** Hana Salehisaheb.

Writing-review & editing: Amirfarhang Miresmaeili, Hana

Salehisaheb.

Competing Interests

The authors declare that they have no conflict of interests.

Ethical Approval

Not applicable.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Greco PM, English JD, Briss BS, Jamieson SA, Kastrop MC, Castelein PT, et al. Banking cases for the American Board of Orthodontics' initial certification examination. Am J Orthod Dentofacial Orthop. 2010;137(5):575-6. doi: 10.1016/j. ajodo.2010.03.003.
- English JD, Briss BS, Jamieson SA, Kastrop MC, Castelein PT, Deleon E Jr, et al. Common errors in preparing for and completing the American Board of Orthodontics clinical examination. Am J Orthod Dentofacial Orthop. 2011;139(1):136-7. doi: 10.1016/j.ajodo.2010.11.005.
- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME). Standards for Educational and Psychological Testing. Washington, DC: AERA, APA, NCME; 1999.
- 4. Nangawe AG. Adoption of web-based assessment in higher learning institutions (HLIs). J. Appl. Res. High. Educ. 2015;7(1):113-22. doi: 10.1108/jarhe-03-2014-0036.
- Komolpis R, Johnson RA. Web-based orthodontic instruction and assessment. J Dent Educ. 2002;66(5):650-8.
- Arevalo CR, Bayne SC, Beeley JA, Brayshaw CJ, Cox MJ, Donaldson NH, et al. Framework for e-learning assessment in dental education: a global model for the future. J Dent Educ. 2013;77(5):564-75.

- Joint Information Systems Committee. Effective Assessment in a Digital Age: A Guide to Technology-Enhanced Assessment and Feedback. JISC Innovation Group; 2010.
- Firestone AR, Beck FM, Beglin FM, Vig KW. Validity of the index of complexity, outcome, and need (ICON) in determining orthodontic treatment need. Angle Orthod. 2002;72(1):15-20. doi: 10.1043/0003-3219(2002)072<0015:votioc>2.0.co;2.
- Savastano NJ Jr, Firestone AR, Beck FM, Vig KW. Validation of the complexity and treatment outcome components of the index of complexity, outcome, and need (ICON). Am J Orthod Dentofacial Orthop. 2003;124(3):244-8. doi: 10.1016/s0889-5406(03)00399-8.
- Cangialosi TJ, Riolo ML, Owens SE Jr, Dykhouse VJ, Moffitt AH, Grubb JE, et al. The ABO discrepancy index: a measure of case complexity. Am J Orthod Dentofacial Orthop. 2004;125(3):270-8. doi: 10.1016/j.ajodo.2004.01.005.
- Schleyer TK, Dasari VR. Computer-based oral health records on the World Wide Web. Quintessence Int. 1999;30(7):451-60.
- Dick RS, Steen EB, Detmer DE. The Computer-Based Patient Record: An Essential Technology for Health Care. National Academies Press; 1997.
- 13. Cozzani M, Weiland F. The European Board of Orthodontists. Int Orthod. 2016;14(2):206-13. doi: 10.1016/j. ortho.2016.03.001.
- 14. Beck CT, Gable RK. Ensuring content validity: an illustration of the process. J Nurs Meas. 2001;9(2):201-15.
- Waltz CF, Strickland OL, Lenz ER. Measurement in Nursing and Health Research. Springer Publishing Company; 2010.
- Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. West J Nurs Res. 2003;25(5):508-18. doi: 10.1177/0193945903252998.
- Daniels C, Richmond S. The development of the index of complexity, outcome and need (ICON). J Orthod. 2000;27(2):149-62. doi: 10.1093/ortho/27.2.149.