Psychometric Properties of a Farsi Version of the Authentic Assessment Perception Questionnaire

Arezoo Ebn Ahmady[1], Shahram Yazdani[2], Azam Valian[3], Zohreh Amiri[4], Fathieh Mortazavi[5], Harry A Lando[6]

[1]Assistant Professor, Community Oral Health Department, Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [2]Professor, Medical Education School, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [3]Azam Valian, Assistant Professor, Department of Operative Dentistry, Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [4]Assistant Professor, Bio-statistic Department, Nutrition Institution, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [5]Medical Education Senior Expert, Medical Education School, Shahid Beheshti University of Medical Sciences, Tehran, Iran [6]Professor, Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA.

Abstract

The term 'authentic assessment' has recently gained widespread use in education. A five-dimensional questionnaire for authentic assessment was translated into Farsi. The questionnaire which comprises 29 items, divided into 5 subscales (task, physical context, social context, result/form and criteria) was developed in English by Dr. Gulikers and her colleagues in the Netherlands. The questionnaire was translated using a forward-backward method and was pilot tested in terms of translation clarity and applicability. The psychometric properties of the Persian version of the questionnaire were evaluated in terms of face, content, and construct validity in addition to test-retest reliability. A convenience sample of 230 dental students (70 males and 160 females) studying in four dental schools in Tehran city was recruited to evaluate the reliability and construct validity of the Persian version. The quality rating of the translations was favorable, suggesting a high quality of both forward and backward translations. The Content Validity Index (CVI) and Ratio (CVR) for the final Farsi version of the questionnaire were found to be acceptable. Cronbach alpha coefficients for all subscales ranged from 0.78-0.91. These preliminary results suggest that a five dimensional questionnaire in its Farsi version may be a valuable tool in dental education assessment and studies.

Key Words: Content Validity Index (CVI); Psychometric properties

Introduction

People can differ in their perception of the authenticity of the same assessment since authenticity is not an objective construct. The problem in this case is that students' perceptions of assessment characteristics appear to have more influence on student learning than do the "objective "characteristics of assessments themselves [1]. If assessment authenticity is defined by the degree of correspondence between the assessment and the professional practice situation it is purported to reflect, then the influence of an authentic assessment on student learning depends on how a student perceives the resemblance between this assessment and professional practice [2].

"Assessment authenticity" is defined as a multidimensional construct with five assessment characteristics [2], namely the assessment task, the physical context of the assessment, the social context of the assessment, the assessment form, and the assessment criteria. The perception of assessment authenticity should be also rated along these five dimensions. The five dimensions questionnaire has been used in different studies of authentic assessment in the Netherlands [3,4]. The measure has proved to be reliable and valid in cross-sectional studies. However, every time a scale is used in a new context or with a different group of people, it is necessary to re-establish its psychometric properties [5]. The objective of this study was to translate and adapt the five dimensions questionnaire of authentic assessment into Persian, the official language of Iran, and to test its reliability and validity with dental students in Iran.

Face validity is the extent to which a test is subjectively viewed as covering the concept it purports to measure. It refers to the transparency or relevance of a test as they appear to test participants. In other words, a test can be said to have face validity if it "looks like" it is going to measure what it is supposed to measure [6,7].

Kerlinger [8] argues that content validity is representative of the content. Thus, content validity of an instrument depends on the adequacy of a specified domain of content that is sampled. Bush [9] pointed out that content validity refers to the degree that the instrument covers the content that it is supposed to measure. It also refers to the adequacy of the sampling of the content that should be measured [10]. Therefore, content validity measures the comprehensiveness and representativeness of the content of a scale.

Construct validity refers to the degree to which inferences can legitimately be made from the operationalizations in your study to the theoretical constructs on which those operationalizations were based'. When working with multiple constructs in a survey study, it is important to satisfy convergent and discriminate validities in order to satisfy construct validity. 'If you can demonstrate that you have evidence for both convergent and discriminate validity, then you've by definition demonstrated that you have evidence for construct validity [11].

Corresponding author: Arezoo Ebn Ahmady, Community Oral Health Department, College of Dentistry, Shahid Beheshti University of Medical Sciences, Daneshjoo Blvd. Evin, Tehran, IR-Iran 1983963113, Tel: +9821-22421813; e-mail: a.ebnahmady@sbmu.ac.ir, aebnahmady@yahoo.com

Material and Methods

Translation and cultural adaptation methods

The translation and cultural adaptation of the five dimensional questionnaire of authentic assessment perception followed Acquardo's multiple approaches [12]. Two initial forward translations were made independently by two translators who were experienced in assessment and education status questionnaires, but were not familiar with the current questionnaire. English was the native language of one; the other was a native speaker of the target language (Farsi). These translations were circulated to the members of the translation panel, who drew up independent proposals. Following this a meeting was organized. In the course of this meeting, all options were reviewed and the primary translation choices were made. Decisions concerning problematic issues were made after the tests on target subjects.

Identifying experts willing to act as panelists to validate the translated questionnaire

A content evaluation panel will normally comprise of the identified domains, or a domain universe in which the judgments are to be made. The panel was therefore selected in accordance with objective criteria dictated by the nature of and required outcomes of the research. Although the Lawshe Method [13] of content validation only requires a minimum of four panelists, it was decided to include as many experts in the panel as practically possible. This further enhanced the value of the model by ensuring that it will be difficult to find many other researchers and practitioners with the credentials or authority to challenge the purported content validity of the model. The specialist nature of the research necessitated that experts dedicate at least half an hour to consider the 5 dimensional frameworks and complete the questionnaire. Due to practical difficulties in involving a large number of experts in such study, it was decided that a maximum of 12 expert panelists would be invited in the judging process. A relatively small group needs to display a relatively high consensus on the validity of the questionnaire and their consensus needs to be reflected in a Content Validity Rate (CVR) value higher than 0.99. Twelve experts were subsequently identified and telephonically invited to participate. Panelists were deemed to be experts for purposes of this research if they possessed at least a PhD, or equivalent. Prior learning was also recognized for purposes of determining the value of a qualification. All panelists must also have been willing to dedicate approximately one hour of their free time to complete the questionnaire. Of the 12 experts who were initially approached, 10 agreed to participate in the research and returned a correctly completed questionnaire. This amounted to a return rate of 83% [14].

Determining the content validity model

Holsti [15] describes content analysis as any technique used for the purpose of making inferences by objectively and systematically identifying specific characteristics of a message.

In this study qualitative evaluation of the questionnaire was also done based on expert feedback receiving necessary corrections regarding grammar, wording, item allocation, and scaling of the items for face validity. Content Validity Ratio (CVR) and the Content Validity Index (CVI) were used to evaluate quantitative content validity of the questionnaire.

In order to determine content validity, Lawshe (1975) and Chadwick et al. (1984) approaches were chosen [13-15]. In the Lawshe model, a questionnaire was developed and structured to guide and allow panelists to indicate clearly their judgment on the essentiality of inclusion of different items in a model. Participant experts were then requested to write the corresponding code in the spaces provided next to each item under the judgment block. The different responses and codes were: E-essential, U-Useful but not essential and N- Not necessary. At the same time a Content Validity Index (CVI) was calculated on a four-part Likert scale for each item based on the mean CVI for relevance, clarity and simplicity of that item, according to the 10 experts' views.

A 24-item questionnaire based on the five-dimensional framework for assessment authenticity perception [2] was used to examine to what degree students perceived the five assessment characteristics (the task, the physical context, the social context, the form, and the criteria) to resemble professional practice. The respective five subscales of the questionnaire were based on a factor analysis of dental students' scores. The items were scored on a five-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree).

The analysis was conducted using the Statistical Package for Social Sciences (SPSS). The cut-off level for statistical significance was taken at 0.05. The internal consistency of the Persian authentic assessment questionnaire was assessed by standardized Cronbach's alpha.

Results

The translated questionnaire with 24 items in 5 dimensions was designed presenting demographic characteristics of dental students. An expert panel was used to determine the content validity of the questionnaire. Evaluation of the face validity of the questionnaire was conducted based on the comments of the expert panel members, including wording, emotional expression, and removing duplicates. Results from this level were checked with two members of the panel who were knowledgeable about the subject and had more active participation in the Feedback forms. The final revised questionnaire had 24 accepted questions.

For content validity purposes, the questionnaire was sent with assessment tables to 10 specialists in the areas of Medical Education, Dental Public Health, Psychology and Instructional Design. These specialists were asked about the Content Validity Ratio (CVR) or the necessity of the subject and Content Validity Index (CVI) or relation to the subject of each question. The CVR minimum acceptable score for each question was considered to be 0.62 and the minimum acceptable score for CVI was equal to 0.79. After analyzing the results, the CVI score for 18 questions was acceptable and for CVR only 6 questions were in the unacceptable range (*Table 1*).

Based on this evaluation, 6 questions of the original items

did not meet pre-assumed criteria and were modified due to lack of minimum content validity based on the judgment of 10 experts and dental students (since we didn't want to delete any item from original questionnaire). Because of the lack of consensus on the first phase of the judgment, in the rest of the study in order to obtain quantitative validation of the questionnaire, again CVR and CVI of each modified item in the questionnaire were calculated based upon another ten experts' and ten students comments and recommendations. Modified items with scores of over 0.62 for CVR and 0.79 for CVI were retained as appropriate ones (*Table 1*).

Confirmatory factor analysis

The questionnaire was distributed after necessary coordination among junior and senior dental students in four available dental schools in Tehran City. They were asked to participate in this study and send us the completed questionnaire. Participation was voluntary and the questionnaires were anonymous. 230 dental students (30.4% male and 69.4% female) completed the questionnaire in a single session along with demographic information sheet (e.g., age and sex).

A Confirmatory Factor Analysis (CFA) with Linear

Table 1. *The result of first and second rounds of Content Validity Index and Ratio's assessments.*

Questions	CVI	CVR	Modification done	CVI after modification	CVR after modification
1	0.86	1	-	-	-
2	0.86	1	-	-	-
3	0.33	0.47	+	0.8	0.73
4	0.81	0.87	-	-	-
5	0.93	0.87	-	-	-
6	0.66	0.60	+	0.86	0.87
7	0.83	0.73	-	-	-
8	0.93	0.87	-	-	-
9	0.74	0.60	+	0.80	0.73
10	0.86	0.87	-	-	-
11	0.70	0.47	+	0.86	0.87
12	0.83	0.73	-	-	-
13	0.86	0.73	-	-	-
14	0.86	0.73	-	-	-
15	0.86	0.87	-	-	-
16	0.93	1	-	-	-
17	0.93	1	-	-	-
18	0.86	0.87	-	-	-
19	0.86	0.73	-	-	-
20	0.86	0.87	-	-	-
21	0.76	0.73	+	0.86	0.87
22	0.73	0.60	+	0.86	0.87
23	0.86	0.87	-	-	-
24	0.93	1	-	-	-

Table 2. *Reliability scores of Cronbach's alpha for the sub-scales of the Authentic Assessment Perception questionnaire in dental students (n=230) in Tehran.*

Sub-scales (Number of questions in each sub-scale)	Task (5)	Physical (4)	Social (4)	Form (4)	Results (7)
Cronbach's alpha	0.87	0.86	0.78	0.84	0.91

To assess the questionnaire's reliability, its internal consistency and homogeneity, the internal consistency reliability coefficient was calculated by Cronbach's alpha. Therefore, questionnaire was given to 230 dental students from different dental schools in Tehran City.

All sub-scales had a reasonable internal consistency, shown in Cronbach's alpha ranging from 0.78 to 0.91 (*Table 2*). Cronbach's alpha for this scale was >0.75 [15]. The test–retest reliability was used to check the stability. Therefore, 40 dental students from Hamedan city in Iran completed the questionnaire twice with an interval of 2 weeks. Pearson Correlations for the test–retest reliability analysis were calculated in each of the scales: Task=0.85, Physical=0.84, Social=0.97, Form=0.86, and result=0.83.

Structural Relationships (LISREL version 8.80, 2006) was applied to examine the construct validity of the questionnaire. LISREL is an iterative software package designed to evaluate the validity of structural equation models that consist of measurement and structural models [16]. A structural model specifies the interrelationship among latent variables. We employed this method to evaluate the construct validity of the 5 dimensions of the Gulikers' questionnaire. Several a priori criteria must be met to establish the best fitting model. Because the Chi-Square statistic is in essence a statistical significance test it is sensitive to sample size which means that the Chi-Square statistic nearly always rejects the model when large samples (more than 200 participants) are used [17,18]. On the other hand, where small samples are used, the Chi-Square statistic lacks power and because of this may

not discriminate between good fitting models and poor fitting models [19]. Due to the restrictiveness of the Model Chi-Square, we used alternative indices to assess model fit such as CFI and RMSEA. The Comparative Fit Index [20] assumes that all latent variables are uncorrelated (null/independence model) and compares the sample covariance matrix with this null model. Values for this statistic range between 0.0 and 1.0 with values closer to 1.0 indicating good fit. A cut-off criterion of CFI ≥ 0.90 was initially advanced. The RMSEA tells us how well the model, with unknown but optimally chosen parameter estimates would fit the populations' covariance matrix [21]. In recent years it has become regarded as 'one of the most informative fit indices' [22] due to its sensitivity to the number of estimated parameters in the model. In other words, the RMSEA favors parsimony in that it will choose the model with the lesser number of parameters. Recommendations for RMSEA cut-off points have been reduced considerably in the last fifteen years. Up until the early nineties, an RMSEA in the range of 0.05 to 0.10 was considered an indication of fair fit and values above 0.10 indicated poor fit [23]. It was then thought that an RMSEA between 0.08 and 0.10 provides a mediocre fit and below 0.08 shows a good fit.

In this study confirmatory factor analysis yielded Chi Sq. = 577.66 with a CFI 0.67, RMSEA 0.08, and P=0.00. While the CFI is lower in this sample of 230 students, the fit indices meet the criteria of Hu and Bentler (i.e., RMSEA< 0.06). Thus, the model was replicated nicely in this sample.

Discussion

This study is a part of a comprehensive assessment. A validated and renewed five-dimensional framework (5DF) for assessment authenticity [3] is used as a tool for describing the 'objective' authenticity of the assessments used in this study and for examining assessment authenticity from the dental student perspective. This framework argues that five assessment characteristics influence the degree of authenticity of the assessment as a whole. The five assessment characteristics can be described as follows:

1. Task. The assessment assignment that defines the content of the assessment

2. Physical context. The environment in which students have to perform the assessment task

3. Social context. The interaction (im) possibilities during the assessment

4. Form. The assessment method, independent of the content

5. Criteria. The characteristics of the performance (product/process) that are of value.

This study is the first to evaluate the suitability of the questionnaire for use among dental students in Iran and the findings revealed that after minor modifications, we had necessary CVI and CVR for the questions and the 5 dimensions model provided an acceptable fit of the data for dental students. The current study was not without limitations. One limitation pertained to the sample itself. Although numerous studies have employed college students as participants and have been well received by the research community, the results of the current study should be treated as tentative until further confirmatory tests of the factor structure of the questionnaire have been conducted using different samples (e.g., for dental students at country level and members of faculty in schools of dentistry).

However, the method and results of the present study are quite strong and it is expected that such findings would generalize quite nicely. The present study also possesses several practical implications and ideas for future study.

In conclusion, the present paper has identified and confirmed a factor structure for the Questionnaire. In doing so, five related dimensions have been identified that support the theoretical notion of authentic assessment. It is hoped that this paper provides researchers and practitioners with a clearer understanding of authentic assessment perceptions and ideas for future investigations.

References

1. Entwistle NJ. Approaches to learning and perceptions of the learning environment, Introduction to the special issue. *Higher Education*. 1991; **22**: 201-204.

2. Gulikers JTM, Bastiaens TJ, Kirschner PA. A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*. 2004; **52**: 67-85.

3. Gulikers JTM, Bastiaens TJ, Kirschner PA. Authentic assessment, student and teacher perceptions: the practical value of the five-dimensional frame work. *Journal of Vocational Education and Training*. 2006 a; **58**: 337-357.

4. Gulikers JTM, Bastiaens TJ, Kirschner PA, Kester L. Relations between student perceptions of assessment authenticity, study approach and learning outcome. *Studies in Educational Evaluation*. 2006 b: **32**: 381-400.

5. Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use (2nd edn.) New York: Oxford University Press; 2000.

6. Weiner IB, Craighead WE. The Corsini Encyclopedia of Psychology. (4thedn) New Jersey: Wiley publishing; 2010. pp. 637-638.

7. Gravetter FJ, Forzano LA. Research Methods for the Behavioral Sciences. (4th edn) California: Wadsworth; 2012. pp. 78.

8. Kerlinger FN. Foundations of behavioral research. (3rd edn.) New York: CBS Publishing; 1986.

9. Bush CT. Nursing research. Virginia: Reston Publishing Company; 1985.

10. Polit DF, Hungler BP. Nursing research principles and methods. (4th edn.) Philadelphia: JB Lippincott Company; 1991.

11. http://www.socialresearchmethods.net/kb/

12. Acquadro C, Conway K, Hareendran A, Aaronson N; European Regulatory Issues and Quality of Life Assessment

(ERIQA) Group. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Health*. 2008; **11**: 509-521.

13. Lawshe CH. A quantitative approach to content validity. *Personnel Psychology*. 1975; **28**: 563-575.

14. Chadwick BA, Bahr HM, Albrecht SL. Social science research methods. USA: Prentice-Hall; 1984.

15. Holsti OR. Content analysis for the social sciences and humanities. Reading, MA: Addison-Wesley Publishing Company; 1961.

16. Jöreskog K, Sörbom D. LISREL 8: User's Reference Guide. Chicago: Scientific Software International Inc; 1996.

17. Bentler PM, Bonnet DC. Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychological Bulletins*. 1980; **88**: 588-606.

18. Bollen KA, Long JS. Testing Structural Equation Models. California: Sage publications; 1993.

19. Kenny DA, McCoach DB. Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling. *Structural Equation Modeling*. 2003; **10**: 333-351.

20. Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletins*. 1990; **107**: 238-246.

21. Byrne BM. Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming. New Jersey: Lawrence Erlbaum Associates; 1998.

22. Diamantopoulos A, Siguaw JA. Introducing LISREL. London: Sage Publications; 1999.

23. MacCallum RC, Browne MW, Sugawara HM. Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*. 1996; **1**: 130-149.