A Review of Articles Concerning Scale Adaptation in the Field of Education [*]

Nagihan Boztunç Öztürk,[1] Melek Gülşah Şahin,[2] Hülya Kelecioğlu [3]

Abstract

This study aims at reviewing the articles written on scale adaptation in the field of education based on certain criteria. A total of 108 articles published in 10 journals indexed in SSCI and/or ULAKBIM during 2005–2014 were analyzed for this study. The "Scale Adaptation Process Check Form" developed by researchers was used in the analysis of the articles.

The results demonstrated that the purpose of the tools of measurement was stated and that the theoretical bases of the structures to be measured were described in the studies of scale adaptation. Conversely, in studies of linguistic equivalence, while the back translation method was normally used, the translated target forms were usually administered only to the target groups. It was concluded that the EFA and CFA were used in combination in most of the studies for determining the construct validity, whereas the criterion validity was not studied generally. The method of internal consistency was usually preferred in calculating the reliability of the scale adapted.

It was observed that the construct validity was generally similar to the one in the original scale, and that when an item was removed from the scale, no item was recommended as replacement for the removed item.

Introduction

When one intends to perform research on a psychological property of individuals, first, the property in question should be measured. For this purpose, tools attaining high levels of validity and reliability should be employed. Either the researchers should develop scales having those features, or they should use any existing scale to measure the properties. Both alternatives have their own difficulties and regulations. Researchers who choose to develop a scale should first get very well familiarized with the property that they wish to measure and they should be able to describe it. Then, they should follow the stages of scale development and develop the scale. Since scale development is a painstaking and tedious process, the latter option—using an existing scale—may look easier at first. If there are any scales measuring the intended property, information such as the following should be

[1] Hacettepe University, Student Affairs, Monitoring ang Evaluation Office, Turkey, nagihanboztunc@hacettepe.edu.tr
[2] Gazi University, Faculty of Education, Department of Educational Sciences, Turkey, gulsah_eroglu@yahoo.com
[3] Hacettepe University, Faculty of Education, Department of Educational Sciences, Turkey, hulyakelecioglu@gmail.com

obtained about the scale: Are the language and the culture in which the original scale has been developed similar to or the same as the native language and culture of the individuals with whom the research is to be conducted? If the answer is "yes," the task of the researcher becomes slightly easier. After that stage, the researcher should find out for what groups the original scale has been developed. If the group of research is similar to the group for which the original scale was developed, the original scale can be used easily. Yet, if the groups are different, the researchers should try using the scale with a group bearing the properties of the group to which the scale will be administered, and they should verify that the scale yields valid and reliable results. For instance, a researcher wishing to measure university students' test anxiety cannot directly use the scale developed for high school students to measure their test anxiety. First, he or she should prove that the scale also measures the test anxiety of university students in a valid and reliable way.

If the answer is "no" to the question regarding the language and culture in which the original scale has been developed, then a process of scale adaptation should be undertaken. The most important problem related to scale adaptation is the development of the original scale in a language and culture different from the language and culture of the target group. In this case, translating the scale from the source language into the language of the target group may not be sufficient. In addition to potential problems stemming from the structure of the two languages, conceptual problems can also be encountered. Some of the concepts and cases in the source language can mean something entirely different in the target language. Thus, Cronbach (1990) points out that individuals may not fully display their capabilities in aptitude tests conceptualized in different cultures. The main reason for this is that a scale developed in a culture does not guarantee the same effect when translated into another language (Sireci and Berberoğlu, 2000).

Researchers may still prefer to adapt a scale rather than develop one. The reasons for choosing to adapt a scale include speed and low cost of this alternative compared to developing one, intercultural comparisons, researchers' lack of sufficient knowledge and skills to develop a scale, the reliability and frequent use of a scale available in literature, and the belief that the results produced in other cultures will be produced in the target group accurately (Hambleton and Patsula, 1999). Despite these, translation and cultural adaptation in particular is of critical importance. The choice of experts to be involved in the translation process is also very important. The process of translation must be conducted with specialists who have good command of both the source and the target languages, who are familiar with both cultures, and who are knowledgeable about the construct measured.

It is pointed out that there are three probable sources of error in the test adaptation process—in other words, there are three factors making adaptation invalid, namely: (a) cultural/linguistic differences, (b) technical issues, designs, and methods, and (3) the interpretation of the results (Hambleton, 2005).

The cultural and linguistic features may differ in terms of the equivalence of the structures, administration of the tests, the format of the items, and students' speed in answering the tests. Biases stemming from the structures measured, the methods employed, and the items used may be available in adaptation work (Van de Vijver and Poortinga, 2005). If the elements of the structure measured are not equivalent in both cultures, this may lead to biases stemming from the structure. Biases stemming from the methods, however, may be the consequence of the properties of the measurement tools and of the implementation process. In the tools of measurement, cultural properties may be mingled with the scores due to the properties unrelated to the structures measured. In the implementation process, however, the communication gap between the givers and takers of the scale and the physical conditions during implementation may be the sources of biases. All these factors must be taken into consideration in the adaptation process, and thus, efforts must be made to eliminate the sources of error.

A second point lowering the validity of adaptation relates to the technical factors. The factors decreasing the validity in this context are the test itself, the choice and education of the translators, the translation process, data collection designs, and data analysis (Hambleton, 2005). The availability of the researchers' prior knowledge of the scale to be adapted prevents potential errors in the adaptation process. The translation of the scales, including culture-dependent expressions, must be performed

very carefully. For instance, in a personality scale, the expression "I like starting the conversation at a party" may not express any significant meaning in cultures where parties for entertainment purposes are not common. Ignoring such expressions leads to errors in the test.

The translation process is one of the most important procedures in the adaptation of a scale. Beyond a mechanical translation from one language into another, the process involves the adaptation of the cultural structure of the scale items and of the structure measured. The completion of this process accurately may reduce the biases stemming from the structure. When a decision is made to adapt a tool of measurement in another language, the following steps in relation to linguistic validity are recommended (Hall, Wilson and Frenkenfield, 2003):

1. Brief and simple language should be used.
2. Translators who are experts in the task and familiar with the topic should be employed.
3. Two translators should be employed. While one of them translates from the original (source) language into the target language, the other should translate back the former translators' text into the source language without seeing the original text.
4. Groups of experts to make corrections to both translations should be used.

These steps are important to achieve the linguistic validity of a scale. Beginning the translation process after exposing the translators to a process of training on test development and on the structure of the scale may contribute to reducing the errors stemming from translation. The designs of forward-translation and back translation are employed in ensuring linguistic validity. In forward-translation, the scale is adapted from the original language into the target language by one or more translators. The adequateness of the translation is checked by a translator or by a group of translators. If necessary, corrections on the translation can be carried out at this stage. The fact that experts can make comparisons between the source language and the target language directly in forward-translation, and that small groups are sufficient for the validity of expert judgements can be listed as the advantages of the technique. Yet, certain hitches are also possible in forward-translation. The fact that the adequateness of a translation is confirmed by translators, that translators can be more competent in one language than the other, or they can be more educated than the monolingual group, and that translators can make guesses about some situations based on their intuitions since they are knowledgeable in the two languages are among the possible hitches of forward-translation (Hambleton, 2005).

Back translation is a widely known and commonly used design. In back translation, the scale is adapted from the source language into the target language by one or more translators. Then, one or more translators adapt the text back from the target language into the source language. The original and the back translated texts are compared, and decisions are made on the equivalence. Efforts are made to ensure the equivalence of both texts (Hambleton, 2005). Back translation assures the confirmation of the semantic equivalence between the source language and the target language; however, the time and cost is greater than forward-translation (Maneesriwongul and Dixon, 2004).

Three designs of data collection are widely used in the process of adaptation. The first is to administer the source form and the target form to individuals who have knowledge of both languages. The data on the validity of the translation can be obtained by comparing the statistics from the scale scores of the same individuals. The second design can be applied when back translation is carried out. The source form and the form translated back are implemented in the group in which the original scale has been developed. Factor analysis is performed for the data collected in this way, and thus, the factor structures of the two forms can be compared. Item equivalence can be ensured by comparing the statistics obtained from the items. The third design is to administer the source form to the source group and the target form to the target group. In this application, the data obtained from the two cultures can be compared by performing the analyses based on the item response theory (Hambleton, 2005).

Having collected the data, the process analysis starts. The analysis must be performed at the level of items and the test. The fit of the item statistics must be examined, and at the level of the test, the compatibility of the test structure with the structure of the source form must first be checked. The evidence for the validity and reliability of the test must be provided. After the completion of the process of analysis, the results must be interpreted by considering the cultural differences.

Scale adaptation necessitates a very careful process of working. The rapid increase of the scale adaptation activities in our country makes us ask the question, "To what extent do the adapted scales undergo those mentioned processes?" The errors in the adaptation process cause errors in the interpretation of the results obtained from the scales. This study examines the adaptation processes that the articles published in the journals reviewed by SSCI and/or ULAKBIM during 2005–2014 went through. The research aims to determine the ways pursued in the adaptation process and to check whether or not the ways are compatible with the process of adaptation by content analysis.

In Turkish educational science literature, there are variety of research articles, conference presentations, and thesis that were examined with content analysis (Arık and Türkmen, 2009; Bektaş, Dündar and Ceylan, 2013; Çiltaş, 2012; Çiltaş, Güler and Sözbilir, 2012; Doğru, Gençosman, Ataalkın and Şeker, 2012; Gökçek, Babacan, Kangal, Çakır and Kül, 2013; Göktaş et al., 2012; Gülbahar and Alper, 2009; Karadağ, 2009; Kılıç Çakmak, Çebi, Mihçi, Günbatar and Akçayir, 2013; Saban, 2009). But only two of them were focused on scale adaptation and scale development (Acar Güvendir and Özer Özkan, 2015; Çüm and Koç, 2013). Unlike the other studies, this study focused only scale adaptation and used 108 articles that has been indexing in SSCI and/or in ULAKBIM.

There are many scale adaptation studies whose results are used very common in other studies. In addition, this study will give a clue for future scale adaptation studies by determining the ways pursued in the adaptation process and to check whether or not the ways are compatible with the process of adaptation.

Method

This research examines the articles concerning scale adaptation published in journals reviewed by ULAKBIM and SSCI in Turkey through the method of content analysis. Content analysis is conducted so as to determine certain words or concepts in a text (Büyüköztürk, Kılıç-Çakmak, Akgün, Karadeniz and Demirel, 2008). The basic aim of content analysis is to arrive at the concepts and relations capable of accounting for the data collected (Yıldırım and Şimşek, 2006). In its broadest sense, content analysis can be considered under three headings: meta-analysis, meta-synthesis (thematic content analysis), and descriptive content analysis (Çalık and Sözbilir, 2014).

This study can be called a descriptive content analysis due to the fact that it involves an analysis and evaluation of the research findings and tendencies of scale adaptation activities performed in the field of education in a descriptive manner (Çalık and Sözbilir, 2014).

Articles Put to Content Analysis

The articles for which scale adaptation activities were analyzed were selected from the journals that had been published in Turkey in the field of education during 2005–2014 (until October 2014), whose full texts were accessed online, and which were indexed in the SSCI and/or in ULAKBIM national database. All of the journals indexed in the SSCI and in the ULAKBIM and which were accessible through the internet were included in the scope of the research. Seven journals indexed in the ULAKBIM national database were selected randomly. The journals included in the research and the distribution of articles according to the journals are shown in Appendix 1.

In selecting the scale adaptation articles published in the journals determined, care was taken to see the phrases "scale adaptation," "Turkish form," "validity and reliability study," and "adaptation into Turkish" in the titles. If an article contained the use of a scale in addition to adapting it, the article was not included in our analyses. Of the 10 journals included in the research, a total of 108 articles carrying those properties were put to content analysis.

A total of 108 articles published in 10 journals during 2005–2014 were analyzed in this study. In the articles analyzed, 52.8% were published in the journals indexed in both SSCI and ULAKBIM whereas 47.2% were published in journals indexed in only ULAKBIM national database. Besides, 30.6% of the articles were published in the 2005–2009 period while 69.4% were published in the 2010–2014 period.

Data Collection Tool

A form called the "Scale Adaptation Process Check Form" was developed by the researchers prior to the content analysis in order not to endanger the validity and reliability of the study. With that form, the researchers aimed to determine the standard criteria for the analysis of the articles. The properties of the scale adaptation process and the problematic cases encountered most frequently in scale adaptation studies were taken into consideration while creating the form.

Five experts of measurement and evaluation were consulted for their opinions with regard to the form in order to evaluate the scope of the form developed. Based on expert opinions, the form was updated and given its final shape. The criteria in the Scale Adaptation Process Check Form were also used in the process of content analysis, and the codings corresponding to the criteria are shown in Table 1.

Table 1. Scale Adaptation Process Check Form

Criteria	Coding
No.	
Name of scale	
Authors	
Journal in which the article was published	
Is the journal included in SSCI ?	(0) No (1) Yes
Year of publication	(1) 2005-2009 (2) 2010-2014
Introduction of the Article	
Stating the purpose of the scale	(0) No (1) Yes
Describing the theoretical bases of the structure measured	(0) No (1) Yes
Introducing the Source Form	
Structure it measures	(0) No (1) Yes
The group to which source form was administered	(0) No (1) Yes
The results of other adaptation studies, if there are any	(0) No (1) Yes
Reliability coefficients	(0) No (1) Yes
Validity coefficients	(0) No (1) Yes
Linguistic Equivalence	
Translation method	(1) Forward Translation (2) Back Translation (3) Undefined
Number of experts included in translations	
Whether or not expert opinions were obtained after translations	(0) No (1) Yes
Whether or not the items of the adaptation form are given	(0) No (1) Yes
Design of data collection for the linguistic equivalence study	(1) Administering the source and target forms to students competent in both languages (2) Administering the target and source forms to monolingual students (3) Administering the source form to the students in the source group, and the target form to the students in the target group (4) Administering the target form to the students in the target group
Analyses	
Studies performed for construct validity	(1)EFA (2) CFA (3) Criterion Group (4) Not performed
Are EFA and CFA performed in the same sample?	(0) No (1) Yes
Criterion Validity	(1) Convergent validity (2) Divergent validity (3) Not performed

Table 1. Continue

Criteria	Coding
Item Analysis	(1) Top-bottom groups analysis
	(2) Item test correlation
	(3) Not performed
Reliability	(1) Internal Consistency
	(2) Test-retest
	(3) Parallel Forms
	(4) Split Half
	(5) Not performed
Findings	
EFA results	(0) The number of dimensions is different from the one in the original scale
	(1) The number of dimensions is same as the one in the original scale
Items in the target form	(1) The same number as in the source
	(2) The number of items is the same but on a different dimension from the dimension in the original scale
	(3) The number of items is the same but corrections were recommended for some items
	(4) An item was removed and a new item was recommended instead
	(5) An item was removed but a new item was not recommended instead
The use of the scale after adaptation	(0) No (1) Yes
The users of the scale after adaptation	(0) The researchers not included in the scale adaptation study
	(1) At least one researcher who was included in the scale adaptation study

Analysis of Data

The five articles randomly chosen from the 108 articles were also analyzed by the three researchers so as to determine reliability in the researchers' codings. The following formula was used in order to determine the consistency between the researchers (Miles and Huberman, 1994):

Reliability= number of reconciliation/number of reconciliation +number of disaccord

By using this formula, the consistency between the researchers was calculated as 0.72.

After the completion of the examination of an article by the researchers, the data were combined. Having formed the whole set of data, the data were checked to see whether or not there were any inconsistencies. If any inconsistencies were observed during content analysis, the article was re-examined by the other researcher. After making the necessary corrections, the percentage and frequency analyses were performed for the data obtained for introducing the research and the adapted scale, the adaptation process, and the use of the adapted scales.

Results

The research findings were presented under such headings as "introducing the research and the adapted scale," "the adaptation process," and "the use of the adapted scales." Stating the tool of measurement, explaining the theoretical bases of the measured structure, and introducing the adapted scale were included in the chapter on introducing the research and the adapted scale; the analysis methods used in the process of scale adaptation and the findings on the final state of the scale were included in the chapter on the process of adaptation. Under the sub-heading, "the use of the scales adapted," the findings on whether or not the scales considered appropriate for implementation by the researchers were used in studies conducted in Turkey were presented.

Introducing the Research and the Adapted Scale

On analyzing the criteria examined within the scope of introducing the research and the adapted scale, it was found that the purpose of the measurement tool was stated in 98.1% of the articles, and that the theoretical foundations of the measured structure were described in 92.6% of

them. In the chapter on introducing the source form, it was found that in 94.4% of the articles, the measured structure of the original articles were described, in 45.4% the group to which the source form was administered was introduced, in 19.4% the results of other adaptation studies were mentioned. On examining the values informing about the validity and reliability of the original scale, it was found that the reliability coefficients were given in 70.4% whereas information about validity was given in 45.4% of the articles.

The Adaptation Process

The final state of the scale obtained in consequence of adaptation addition to the procedures related to validity and reliability were addressed in the adaptation process. Hence, in the adaptation process, the findings were listed under the headings of linguistic equivalence, validity, reliability, and the items of the adapted scale.

1. Linguistic Equivalence

Primarily, the methods used in translating the original scale into Turkish were analyzed in the process of determining the linguistic equivalence. The translation methods were handled as forward translation and back-translation. The findings obtained are shown in Table 2.

Table 2. Use of Translation Methods

Translation Methods	Frequencies (f)	Percentages (%)
Forward Translation	33	30.6
Back Translation	75	69.4
Total	108	100

According to Table 2, it was observed that the method of back-translation was used in 69.4% of the adaptation studies analyzed. It was also found that in 30.6% of the studies, only forward translation was used. Accordingly, the method of back translation was preferred more in the translation process. The number of experts used in translation was seen to range between 1 and 25. Besides, in 77.8% of the studies, expert opinion was obtained for the adapted scales. In 54.6% of the articles analyzed, it was found that the items of the scale translated from the original into the target language were given.

The findings on the designs used in the analysis of whether or not linguistic equivalence was attained are shown in Table 3.

Table 3. Design of Data Collection Used in Testing the Linguistic Equivalence

Testing the Linguistic Equivalence	Frequencies (f)	Percentages (%)
Administering the target form to the students in the target group	45	41.7
Administering the source and target forms to students competent in both languages	37	34.3
First and second designs were administered to two different groups	9	8.3
Undefined	17	15.8
Total	108	100

According to Table 3, it was found that only the scale translated into Turkish was administered to the target group in 41.7% of the adaptation articles. In other words, the intelligibility of the scales was examined only in terms of Turkish in those studies. In 34.3%, however, both the original scale and the scale translated into Turkish were administered. Besides, 8.3% of the studies were found to implement these two designs of data collection separately. On examining the size of the groups of implementation in the process of determining linguistic equivalence, it was found that the size of groups ranged between 94 and 3485.

2. Validity

The methods used primarily in determining the construct validity were examined in the process of determining validity. The findings are shown in Table 4.

Table 4. Distribution of the Methods Used in Determining the Construct Validity

Methods Used in Determining the Construct Validity	Frequencies (f)	Percentages (%)
Exploratory Factor Analysis	21	19.4
Confirmatory Factor Analysis	20	18.5
Exploratory and Confirmatory Factor Analyses	56	51.9
Exploratory Factor Analysis and Criterion Groups	2	1.9
Confirmatory Factor Analysis and Criterion Groups	1	0.9
Exploratory Factor Analysis, Confirmatory Factor Analysis and Criterion Groups	5	4.6
Not performed	3	2.8
Total	108	100

According to Table 4, Exploratory and Confirmatory Factor Analyses were used in combination in 51.9% of the adaptation studies. In 19.4% of the studies, only EFA was used, while in 18.5%, only CFA was used. In 4.6% of the studies, EFA, CFA, and criterion groups were used in combination. In 2.8%, however, construct validity studies were not performed. Accordingly, it was concluded that the EFA and the CFA—the methods of determining construct validity—were frequently used in combination.

In studies in which EFA and CFA were used in combination, whether the groups in which those analyses were performed were the same or different was examined separately, according to the 2005–2009 and the 2010–2014 periods. The findings are shown in Table 5.

Table 5. Groups in Which EFA and CFA were Used

Year of Publication of the Article	Are EFA and CFA tested in the same group?	Frequencies (f)	Percentages (%)
2005-2009	Yes	11	100.0
	No	0	0
2010-2014	Yes	42	84.0
	No	8	16.0
Total		61	100.0

According to Table 5, EFA and CFA were used in the same groups in all of the articles published in the 2005–2009 period while they were used in different groups in 16.0% of the articles. Thus, it was found that the EFA and CFA were starting to be preferred with different groups from 2005 to 2014.

The methods used in determining the criterion validity, which is another type of validity, are shown in Table 6.

Table 6. Distribution of the Methods Used in Determining the Criterion Validity

Methods Used in Determining the Criterion Validity	Frequencies (f)	Percentages (%)
Convergent validity	28	26.0
Divergent validity	0	0.0
Convergent and Divergent validity	13	12.0
Not performed	67	62.0
Total	108	100

According to Table 6, criterion validity was not tested in 62.0% of the studies. In studies where criterion validity was performed, however, convergent validity was preferred more. The studies in which convergent and divergent validity were used together constituted 12.0% of the articles examined. Based on these findings, it was found that criterion validity was not often determined in the scale adaptation process.

The methods used in determining item analysis in the articles examined are shown in Table 7.

Table 7. Distribution of the Methods Used in Item Analysis

Methods Used in Item Analysis	Frequencies (f)	Percentages (%)
Top-Bottom Group Analysis	4	3.7
Item-Test Correlation	42	38.9
Top-Bottom Group Analysis and Item-Test Correlation	31	28.7
Not performed	31	28.7
Total	108	100

According to Table 7, only the item-test correlation was examined in 38.9% of the articles. In approximately one fourth (28.7%) of them, item analysis was not performed. Top-bottom group analysis and item-test correlations were conducted together in 28.7% of the studies.

3. Reliability

The methods used in predicting reliability were evaluated, and the findings obtained are shown in Table 8.

Table 8. Distribution of the Methods Used in Predicting Reliability

Methods Used in Predicting Reliability	Frequencies (f)	Percentages (%)
Internal Consistency	50	46.3
Internal Consistency and Test-retest	35	32.4
Internal Consistency and Parallel Forms	3	2.8
Internal Consistency and Split Half	6	5.6
Internal Consistency, Test-retest, and Split Half	13	12.0
Internal Consistency, Parallel Forms, and Split Half	1	0.9
Total	108	100

According to Table 8, the method of internal consistency was preferred in predicting the reliability in 46.3% of the studies examined. In 32.4% of the articles, internal consistency method was used in combination with the test-retest method. 12.0%, of the studies were found to use internal consistency, test-retest, and split half methods in combination. Therefore, internal consistency may be said to occupy an important place in scale adaptation activities. Almost all of the studies calculated the Cronbach α coefficient in order to predict the reliability in terms of internal consistency.

The items of the scales formed in consequence of the scale adaptation process were examined, and the findings obtained are shown in Table 9.

Table 9. Properties of the Items of the Scale Formed in Consequence of Adaptation

Items of the Scale Formed in Consequence of Adaptation	Frequencies (f)	Percentages (%)
The same number as in the source	66	61.1
The number of items is the same but on a different dimension from the dimension in the original scale	4	3.7
An item was removed but a new item was not recommended instead	38	35.2
Total	108	100

According to Table 9, the number of items in the original scale was the same as in the adapted one in 61.1% of the studies. In 35.2% of them, however, items were removed from the original scale, but no items were recommended instead. In 3.7%, the adapted scale had the same number of items as the original one, but the items were on different dimensions from the ones in the original scale. Accordingly, it may be said that most of the adapted scales have the original scale. Yet, there were also considerable amount of scales with removed items (35.2%). It was found that researchers did not recommend items for that situation, which damaged the structure of the scale.

The Use of the Adapted Scales
The use of the adapted scales in later research studies is as important as adaptation activities. Therefore, whether or not all of the scales included in this research and published in the 2005–2009 period, and chosen 25 of those published in the 2010–2014 period were used in later studies (for instance in articles, presentation) was examined here. While the analysis was being performed, information on the title of the scale, the author's name and surname, and the year of publication of the scale were searched through Google, and they were used as the key words. The findings obtained are shown in Table 10.

Table 10. Whether or not the Adapted Scale was used in Later Studies

Are the adapted scales used in later research studies?		Frequencies (f)	Percentages (%)
2005–2009	No	7	21.2
	Yes	26	78.8
2010–2014	No	17	68.0
	Yes	8	32.0
Total		58	100

According to Table 10, 78.8% of the scales belonging to the 2005–2009 period were used in later research studies. Besides, 32.0% of the 25 scales chosen from the ones belonging to the 2010–2014 period were used in later research studies. It may be said that the reason for the lower use of the scales in the 2010–2014 period is that the date of their publication is quite recent.

The findings concerning the users of the scales in later research studies are shown in Table 11.

Table 11. Users of the Scales in Later Research

Who used the scales in later research studies?		Frequencies (f)	Percentages (%)
2005-2009	The research group composed of the researchers not included in the adaptation activity	22	84.6
	The research group containing at least one of the researchers included in the adaptation activity	4	15.4
2010-2014	The research group composed of the researchers not included in the adaptation activity	7	87.5
	The research group containing at least one of the researchers included in the adaptation activity	1	12.5
Total		34	100

As is clear from Table 11, 84.6% of the scales belonging to the 2005–2009 period, and 87.5% of the scales belonging to the 2010–2014 period were used by the researchers not included in the adaptation activities. In consequence, the scales adapted were mostly used by different researchers. This is the same for both periods of five years.

Conclusion and Suggestions

Based on the research findings, it may be said that the processes to follow in scale adaptation were taken into consideration in the majority of the articles examined. This is important for the use of adapted scales in different studies because the measurement results will be different from the intended measurement results unless the adaptation stages are followed appropriately.

It is remarkable that criterion validity was less preferred in analyzing the validity of the scales. Yet, criterion validity presents significant information about the extent to which the adapted scale is related to the criterion for which reliability and validity were performed. Considering the fact that construct and criterion validity can be prioritized in studies that measure affective properties such as attitudes (Büyüköztürk et al., 2008), the importance of criterion validity can be clearly understood.

In general, only the item-total test score correlations were calculated in the analysis of the scale items. In studies where criterion validity was performed, however, convergent validity was preferred more frequently. It was observed in examining the reliability of the scales that the method of internal consistency was used in all of them. While reliability was analyzed only in terms of internal consistency in some of the scales, reliability was analyzed both in terms of internal consistency and stability in others. In almost all of the articles, Cronbach α coefficient was preferred in calculating the internal consistency, whereas in some others, Spear Brown's split half in addition to Cronbach α was analyzed.

It was seen that more than half of the scales published in the 2005–2009 period were used in later research studies. Besides, the majority of those articles were used by the researchers who were not included in adaptation activities. On examining the group of articles selected from the ones published in the 2010–2014 period, it was found that more than half of the scales were not used in a different research study. This may have stemmed from the fact that the adapted scales were published very recently. It was observed that more than half of the selected 25 scales were used by the researchers who were not included in the adaptation activities. This can be regarded as the indicator of the fact that adaptation activities have attained their objectives.

In some of the adaptation activities, it was observed that some items were removed, but that no items were recommended instead. The decision to remove items was generally made on the basis of statistical results only. When the items are removed on the basis of statistical results only, significant changes can occur in the structure of the scale. In this case, the scores obtained in the scale

may not mean the same as in the original scale. If an item is to be removed from the scale in consequence of the adaptation process, the researchers developing the original scale and the subject matter experts should be contacted to obtain better results because this case also modifies the structure of the scale. Prior to removing items, researchers should discuss with the developers of the scale and with the experts knowledgeable in the structure of the scale and in both cultures. The decisions in relation to statistically inconsistent items should be made after performing detailed analyses based on the structure of the scale in both cultures. The decisions made should be justified by relating them to the structure of the scale and to the cultures.

It was concluded that due importance was not generally attached to criterion validity in adaptation activities. Criterion validity plays an active role in the process of scale adaptation—where the scales are tested with similar or different structures and an external source is used—and it also presents information on the structure of the scale. Therefore, due importance should be attached to criterion validity in the scale adaptation process.

Standardization may be introduced for the scale adaptation process by considering the criteria used in this research. The standards to be developed will both guide the researchers who perform adaptation activities and will assure that the activities are evaluated according to standards.

Scale adaptation activities published in foreign journals with various effects can be examined based on those criteria. Those can be compared with the ones published in Turkey. Adaptation activities performed in Turkey and abroad can also be compared according to years.

Similar research studies can also be conducted for scale development activities. Thus, standards can also be set in the process of scale development.

References

Acar Güvendir, M., & Özer Özkan, Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi, 14*(52), 23-33.

Arık, R. S., & Türkmen, M. (2009, Mayıs). *Eğitim bilimleri alanında yayımlanan bilimsel dergilerde yer alan makalelerin incelenmesi.* I. Uluslararası Türkiye Eğitim Araştırmaları Kongresi'nde sunulan bildiri. Çanakkale Onsekiz Mart Üniversitesi, Çanakkale.

Bektaş, M., Dündar, H., & Ceylan, A. (2013). Ulusal sınıf öğretmenliği eğitimi sempozyumu (USOS) bildirilerinin çeşitli değişkenler açısından incelenmesi. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, *6*(2), 201-226.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2008). *Bilimsel araştırma yöntemleri.* Ankara: Pegem Akademi Yayınları.

Cronbach, L. J. (1990). *Essentials of psychological testing*. New York: Harper&Row Publishers.

Çalık, M., & Sözbilir, M. (2014). İçerik analizinin parametreleri. *Eğitim ve Bilim*, *39*(174), 33-38. doi: 10.15390/EB.2014.3412

Çiltaş, A. (2012). 2005-2010 yılları arasında matematik eğitimi alanında Türkiye'de yapılan yüksek lisans ve doktora tez çalışmalarının içerik analizi. *The Journal of Academic Social Science Studies*, *5*(7), 211-228.

Çiltaş, A., Güler, G., & Sözbilir, M. (2012).Türkiye'de matematik eğitimi araştırmaları: bir içerik analizi çalışması. *Kuram ve Uygulamada Eğitim Bilimleri*, *12*(1) ,565-580.

Çüm, S., & Koç, N. (2013). Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi, *Eğitim Bilimleri ve Uygulama, 12*(24), 115-135.

Doğru, M., Gençosman, T., Ataalkın, A. N., & Şeker, F. (2012).Fen bilimleri eğitiminde çalışılan yüksek lisans ve doktora tezlerinin analizi. *Türk Fen Eğitimi Dergisi*, *9*(1), 49-64.

Gökçek, T., Babacan, F. Z., Kangal, E., Çakır, N., & Kül, Y. (2013). 2003-2012 yılları arasında Türkiye'de karma araştırma yöntemiyle yapılan eğitim çalışmalarının analizi. *International Journal of Social Science Studies*, *6*(7), 435-456. doi: 10.9761/JASSS1655.

Göktaş, Y., Küçük, S., Aydemir, M., Telli, E., Arpacık, Ö., Yıldırım, G., & Reisoğlu, İ. (2012). Educational technology research trends in Turkey: a content analysis of the 2000-2009 decade. *Educational Sciences: Theory & Practice*, *12*(1), 191-196.

Gülbahar, Y., & Alper, A. (2009). Trends and issues in educational technologies: A review of recent research in TOJET. *TheTurkish Online Journal of Educational Technology – TOJET*, *8*(2), Article 12.

Hall, E. O. C., Wilson, M. E., & Frankenfield, J. A. (2003). Translation and restandardization of an instrument: The early infant temperament questionnaire. *Journal of Advanced Nursing*, *42*(2), 159-168.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1*(1), 1-30.

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 3-38). Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers.

Karadağ, E. (2009). Eğitim bilimleri alanında yapılmış doktora tezlerinin incelenmesi. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, *10*(3), 75-87.

Kılıç Çakmak, E., Çebi, A., Mihçi, P., Günbatar, M. S., & Akçayir, M. (2013). A content analysis of educational technology research in 2011. 4th International Conference on New Horizons in Education. INTE 2013 Proceedings Book, 397-409.

Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: a methods review. *Journal of Advanced Nursing*, *48*(2), 175-186.

Miles, M. B., & Huberman, A. M. (1994). *An expanded sourcebook qualitative data analysis*. London: SAGE Publications.

Saban, A. (2009). Çoklu zekâ kuramı ile ilgili Türkçe çalışmaların içerik analizi. *Kuram ve Uygulamada Eğitim Bilimleri, 9*(2), 833-876.

Sireci, S. G., & Berberoğlu, G. (2000). Using bilingual respondents to evaluate translated-adapted items. *Applied Measurement in Education*, *13*(3), 229-248.

van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 39-63). Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers.

Appendix 1. Journals Included in the Research and the distribution of the Articles according to the Journals

Journal	Index	Number of Articles
Education and Science Journal	SSCI/ULAKBIM	23
Educational Sciences: Theory & Practice	SSCI/ULAKBIM	21
Hacettepe University Journal of Education	SSCI/ULAKBIM	14
Ankara University Journal of Educational Sciences	ULAKBIM	12
Elementary Education Online	ULAKBIM	10
Journal of Kırşehir Education Faculty	ULAKBIM	8
Kastamonu University Kastamonu Education Journal	ULAKBIM	7
Mersin University Journal of the Faculty of Education	ULAKBIM	7
İnönü University Journal of the Faculty of Education	ULAKBIM	3
Journal of Uludağ University Faculty of Education	ULAKBIM	3
Total		108