

An Analysis of Scale Adaptation Studies in Science Education: Meta-Synthesis Study

Behiye Akcay, Burcu Gelen, Aydin Tiryaki, Ibrahim Benek
Istanbul University-Cerrahpasa

To cite this article:

Akcay, B., Gelen, B., Tiryaki, A. & Benek, I. (2018). An analysis of scale adaptation studies in science education: Meta-synthesis study. *Journal of Education in Science, Environment and Health (JESEH)*, 4(2), 227-245. DOI:10.21891/jeseh.439150

This article may be used for research, teaching, and private study purposes.

Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles.

The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material.

An Analysis of Scale Adaptation Studies in Science Education: Meta-Synthesis Study

Behiye Akcay, Burcu Gelen, Aydin Tiryaki, Ibrahim Benek

Article Info

Article History

Received:
29 April 2018

Accepted:
28 June 2018

Keywords

Scale
Adaptation
Science education
Meta-synthesis
Content analysis

Abstract

The purpose of this study was to examine scale adaptation studies in the field of science education in terms of content, method as well as to determine the general guidelines used in the scale adaptation process. The population of this study included a total of 145 journals published in the field of science education in Turkey. A total of 46 adaptation studies which were published in 25 journals were selected from this population by using the purposive sampling methods comprise of the study sample. Meta-synthesis method was used for the data analysis in this study. Results of the study showed that adaptation studies in Turkey were mostly conducted in the self-efficacy area. It was determined that all studies calculated the internal consistency coefficient and that likert-type scales were commonly used. This study also determined that the language validity stage was not sufficiently discussed and was only briefly mentioned in the examined adaptation studies. It was found that none of the studies translated original test guidelines to target language. This study concluded that all of the adapted studies did not get official permissions for the pilot study. In addition, it was found that the researchers preferred to use CFA (Confirmatory Factor Analysis) and EFA (Exploratory Factor Analysis) methods together for the construct validity.

Introduction

There are many assessment tools in social sciences to measure a variable. However, it is important to find the most reliable and valid tool for the study objective (Atilgan 2015). This scale can either be developed when needed or adapted from another language. Secer (2015) described scale adaption studies as the use of a valid and reliable scale, which was developed in a foreign language, in a different language and culture by conducting its validity and reliability analyses.

In the relevant literature, the first scale adaptation was made by Lewis M. Terman in 1916. Terman adapted the Intelligence Test, which was developed in French by Alfred Binet and Theodore Simon in 1905, and changed the name to the Measurement of Intelligence for the American Culture (Tsagari and Floros, 2013). Nowadays, there is a growing need to compare the success of students from different countries (Deniz 2007; Sireci 1995). One of the main goals of comparing is to share educational practice and promote growth and improvement, i.e., to learn from each other. There have been many studies comparing students' academic success in different cultures since 1970s. In 1995, for instance, 40 countries participated in the Third International Mathematics and Science Study, and the scale was adapted to 30 languages (de Vijver and Hambleton, 1996). In 2015, fourth- and eighth-grade students from 57 countries participated in the Trends in International Mathematics and Science Study (TIMSS), and the test was adapted to 43 languages (Deniz 2007; Ebbs and Korsnakova, 2015; Polat, Gonen, Parlak, Yildirim, and Ozgurluk, 2015). The Programme for International Student Assessment (PISA) is one of the internationally accepted largest-scale of standardized assessment tool for 15-year olds. PISA was used for 72 countries in 2015 to evaluate students' skills and knowledge (the Organisation for Economic Co-operation and Development [OECD] 2017).

In the *Standards for Educational and Psychological Tests and Manuals*, which was revised by the American Psychological Association (APA), American Educational Research Association (AERA) and National Council on Measurement in Education (NCME) in 2014, it states that the main purpose for developing "Educational and Psychological Tests" is to develop a universal method that is accessible to researchers from all countries (AERA, APA and NCME 2014).

The International Test Commission (ITC) emphasized the importance of language translations made in scale adaptations in 1993, and published their first draft report in 1994. The commission completed their report in 1996, and explained 22 binding rules in detail (de Vijver and Hambleton 1996). The ITC classified these rules, which were developed for scale adaptation, under five titles (de Vijver and Hambleton, 1996; Hambleton, Merenda and Speilberger 2009): (1) context, (2) test development and adaptation, (3) application of test, (4) documentation/score evaluation, and (5) general and professional conditions.

Within the scope of *context*, the effect of cultural differences should be minimized as much as possible. The selected sample and the sample to which the original sample is applied should be in accordance with each other (de Vijver and Hambleton 1996; Plake and Wise 2014). During the *test development and adaptation* stage, researcher(s) should consider that this test may be adapted to a different culture and/or language in the future. Test developers should prepare directions, a scoring table and items of the test considering the cultural and linguistic characteristics of the sample. The selected test method, item format and other procedures should comply with the target group. They should also ensure that item content is suitable for the target group. They should test whether the scale, which is adapted in linguistic and psychological aspects, complies with the original scale. Appropriate statistical methods should be used. At the end of the statistical analysis, conflicting items should be excluded from the test. Validity and reliability of the adapted scale should be ensured, and according to analyses results, items determined to be inappropriate for the target group should be excluded from the test (Hambleton, Merenda and Speilberger 2009; Plake and Wise 2014). In the *test application* stage, the selected framework should be similar to the original scale. In this stage, possible problems that might emerge during the use of the scale should be determined and solved in advance. Researchers should also review the materials and application procedures that may affect the validity of the test. Test directions in the original and translated scales should be similar to each other. Test booklets should be suitable for the target culture. Rules to be followed during the application stage should be specified clearly. If the test is applied to a sample group that is different than the original, the reason for this change should be specified in the *documentation/score evaluation* stage. The researcher is responsible for score differences that may occur between the samples. In the event that the scores obtained from the sample comparison are similar, results should be reported. The researcher should state that socio-cultural and ecological effects on the test results can affect the performance of the sample group, and also state how the results can be affected. In context of *general and professional conditions*, test developers should be aware of the fact that linguistic and cultural differences of the target group are important in test adaptation studies (Hambleton, Merenda and Speilberger 2009).

The Advantages and Disadvantages of Scale Adaptation Studies

The negative aspects of scale adaptation:

- There may be difficulty in ensuring cultural and linguistic equivalency (Cha, Kim and Erlen 2007; Grisay, 2000; Secer 2015, Seker and Gencdogan 2006),
- There is a possibility that a number of items and factors in the original version change as a result of adaptation process (Secer 2015; Seker and Gencdogan 2006),
- If original version of the scale has more than five-point Likert-type options, it is not suitable for Turkish language (de Vijver and Hambleton 1996; Secer 2015; Seker and Gencdogan 2006).

The positive aspects of scale adaptation:

- It is convenient for researchers in terms of time, effort and economic aspects (Cha, Kim and Erlen 2007; Savasir 1994) since the examination of the theoretical background is a time-consuming, difficult and expensive process (Deniz 2007).
- It is probable that the researcher does not have the technical information related to the subject. In this case, adaptation of the scale is more suitable (Deniz 2007; Savasir 1994; Secer 2015).
- Adaptation studies allow researchers to make comparisons between different cultures (Deniz 2007; Hambleton et al. 2009; Secer 2015; Seker and Gencdogan 2006).
- In countries where different languages are spoken, similar versions of the same scale that are developed for the spoken languages will enable analyzing study results in a more efficient way (Savasir 1994).
- An adaptation of a widely-known and commonly-used scale will be trusted more than a newly developed test (Deniz 2007).

The Primary Stages Followed in Scale Adaptation Studies

Preparation Stage: In this stage, the researcher determines the need for scale adaptation (Akbas and Korkmaz 2007), decides which scale to translate, obtains the permissions for the adaptation study, and forms the translation team (Gudmundsson 2009; Secer 2013).

Language Validity: This stage includes translating the scale items and directions from the source language to the target language, examining and comparing these translations, making a backward translation, giving its first form to the translated scale, performing an application for language validity, performing statistical analyses for language validity, and giving its form to the scale of which the language validity is constructed (Geisinger 1994; Gudmundsson 2009; Savasir 1994; Seker and Gencdogan 2006).

Ensuring cultural and linguistic equivalency is the most important process in scale adaptation studies (Cha et al. 2007; Deniz 2007; Secer 2015; Savasir 1994) because the translation team including field experts having a good command of both languages is the most efficient factor to make the adaptation study successful. According to the Standards for Educational and Psychological Testing which was developed from AERA, APA and NCME the validity and reliability of a test should be tested by applying it to a group comprising of linguists when it is translated from one language to another language or dialect (Hambleton 2009; Tsagari and Floros 2013).

According to Tsagari and Floros (2013), popular translation methods in test adaptations are forward translation, which means translating test from the source language to the target language, and backward translation, which means translating the scale –which is already translated to the target language– back to the source language (Grisay 2000; Savasir 1994). Cha, Kim and Erlen (2007) stated that backward translation is the most important step in scale adaptation studies. Independent linguists should work on the translation in these two translation processes. However, Tsagari and Floros (2013) believe that these adaptation studies are challenging in any circumstances since there is a possibility for the linguists who make the translation to make subjective decisions, and ignore some points even if they speak the source language. Therefore, Maxwell (1996) and Savasir (1994) stated that a good translator should have an excellent knowledge of both the source and adaptation languages, have enough experience to recognize the cultural structures of both languages as well as the sample group to which the scale is to be adapted, and skills for test development.

Pilot Study: This stage comprises of obtaining official permissions for a pilot study, carrying out the pilot study, and, performing validity and reliability analyses after the pilot study (Geisinger 1994; Gudmundsson 2009; Savasir 1994; Seker and Gencdogan 2006).

It is required to obtain official permissions for the sample group to which the application will be performed in order to conduct the pilot study of the adapted form of the scale. After receiving the necessary permissions, the pilot study is conducted and the obtained data are analyzed using the SPSS, AMOS or LISREL software. To determine the sample size, the Kaiser-Meyer-Olkin (KMO) value is calculated. KMO values higher than 0.60 show that the sample size is suitable (Buyukozturk 2009). Barlett's test is used to determine whether data are suitable for factor analysis, and they will be if the test result is significant (Secer 2013). For item analysis of the test, differences between items mean scores of 27% sub above groups are calculated using an unrelated t-test (Buyukozturk 2009).

For the validity analyses; content, structure and criterion validity are calculated. Content validity is calculated in the original form; thus, it is not necessary to calculate it again in adaptation studies. In the content validity step, a table of specifications is created for matching between target behaviors and items of the measurement tool. Field experts examine the table of specifications and report their opinions for the content validity. The inter-rater reliability level is expected to be 90-100% (Buyukozturk 2009; Turgut and Baykul 2012). Criterion validity is divided into two: concurrent validity and predictive validity (Turgut and Baykul 2012). Two types of factor analyses are performed for construct validity of the scale in the measurement tool development and adaptation stage:

Exploratory Factor Analysis (EFA): This is generally used for scale development. Its purpose is to determine the implicit structure of the scale (Secer 2015). This analysis is performed to test the appropriateness of scale items to the predefined theoretical structure. It is used to determine the number of sub dimensions in which the scale items can be included. Moreover, this analysis is important for the determination of the relationship between scale items (Secer 2013).

Confirmatory Factor Analysis (CFA): This analysis is used for scale adaptation and scale development studies. Its purpose is to confirm the implicit structure which is determined using the EFA. It is used to determine which structures are identified by a series of items. In scale adaptation studies, it ensures the control of the model fit between factorial structures of the original version and the translated version (Secer 2015). While adapting a scale, some researchers prefer to perform a CFA only rather than performing both analyses (EFA and CFA) together. However, It is necessary to perform the CFA to determine whether there is a difference in the factorial structure of the two scales, if it is to be adequately adapted to a new language and culture. In addition to enabling researchers to test the hypotheses for factor structure and model fit, it provides more informative analytic options.

For *reliability* calculations; internal consistency coefficient, split-half reliability, and test-retest calculations are conducted. Reliability is used in the meaning of consistency during developing and adapting a measurement tool. Consistency is receiving similar or the same results when measurements are repeated (Buyukozturk 2009). Kuder Richardson-20 (KR-20) and Cronbach's alpha internal consistency value (Atilgan 2015) are used as evidence for the content validity. The KR-20 is used in the scoring of two categories. KR-20 statistics are calculated when variance is determined. Item difficulty statistics should be known for this calculation. It is performed in one-dimension scales. If there are raw scores, then it is calculated. The Cronbach's alpha reliability coefficient is calculated in likert-type scales with three or more points (Gomleksiz and Erkan 2010).

As reliability estimation techniques; test-retest, equivalent forms, and test-split test methods are preferred. The *test-retest method* is explained with the Pearson product-moment correlation which is obtained as a result of administering a test to the same group periodically (Atilgan 2015). For the *equivalent (alternative/parallel) forms method* two equivalent forms, which are developed to measure the same characteristic, are administered to the same group at the same or different times, and the consistency between scales are calculated by applying the Pearson product-moment correlation on the results (Buyukozturk 2009). The *test-split test method (equivalent splits – split-half)*, is important for calculating the consistency between test scores. It is appropriate for one-dimension tests. Items in the test split in equivalent halves as odd-even or first half-last half (Atilgan 2015). The Pearson product-moment correlation is calculated using scores obtained from both halves of the test. The obtained value is accepted as the reliability of either one of the halves. The reliability for the entire scale is calculated using the Spearman-Brown formula (Atilgan 2015). If test items show a heterogeneous distribution in terms of the measured characteristic, it will be appropriate to determine reliability using the test-split test method rather than using an internal consistency coefficient (Gomleksiz and Erkan 2010).

The Studies Conducted in Turkey

By using thematic content analysis (meta-synthesis), Gul and Sozbilir (2015) analyzed 22 scale-development studies in science and mathematics education field, which were published between the years 2000 and 2013 in six primary academic journals in Turkey focusing on educational sciences, and selected based on the pre-determined criteria. They found that attitude scale was the most commonly developed type of scale, and these studies were mainly conducted in mathematics education field. Of the validity analyses, researchers mainly worked on construct validity, while confirmatory and exploratory analyses were studied at a moderate level. Gul and Sozbilir (2015) explained that they analyzed validity of the 22 scale-development articles based on criterion validity which was found only in one article, content validity which was calculated more than half of the articles, face validity was calculated only in five articles. In addition in terms of construct validity the researchers found that only eight articles used discriminant validity while there was no information about convergent validity. It was also determined that internal consistency methods were the favorite of researchers among reliability studies.

Guvendir and Ozkan (2015) examined studies which were conducted on scale development and adaptation in Turkey between the years of 2006 and 2014, and they aimed to determine stages used in scale development and adaptation processes. The researchers selected a total of 59 articles on scale adaptation and development, and these studies comprised the study sample. The study found that two or more translators worked in the adaptation studies, and the directions of the original scale were not translated in any of the articles. Only one article consulted assessment experts, and few studies calculated equivalent split-half reliability.

Boztunc-Ozturk, Eroglu and Kelecioglu (2015) examined scale adaptation studies which were conducted in the education field between the years of 2005 and 2014. A total of 108 article studies matching the determined criteria comprised the study sample. In the data analysis, the study used the “Control Form for the Scale Adaptation Process” form developed by the researchers. According to the study results, all studies used the

backward translation method, and the language equivalency calculation used only the adapted form, while the original scale was used in none of the studies. For construct validity, an EFA and CFA were used together. None of the studies examined the criterion validity of the scale. As an internal consistency coefficient, the Cronbach's alpha value was calculated. According to the construct validity results, the adapted scales showed similarity with the original scale, and no new items were suggested to replace the excluded items.

Delice and Ergene (2015) examined scale development and adaptation studies, which were conducted in Turkey in mathematics education field between the years of 2005 and 2014, according to certain criteria. The study sample included 35 scale development and 18 scale adaptation studies. That study found that only two-thirds of these studies mentioned whether there was a need for a scale adaptation study. More than half of these studies did not mention that necessary permissions were obtained for the adaptation of the scale. The researchers made certain that the translation team created for the adaptation included only individuals that spoke the target language. It was determined that researchers were not careful enough to form their groups of experts that were sufficiently informed about both source and target languages and cultures as well as the characteristics of the scale to be adapted. All studies calculated the Cronbach's alpha value. It was found that the calculated Cronbach's alpha values were in compliance with the original scales. Some studies did not report their reliability analysis results (internal consistency coefficients). It was also reported that the sample size used in some studies was not big enough, while Delice and Ergene (2015) suggested that five to 25 subjects need for per variable. Costello and Osborne (2005) suggest a ratio of 10/1 as a minimum but recommend a ratio of 20/1 subject to variable" (p.7).

Cum and Koc (2013) examined scale development and adaptation studies which were conducted in the psychology and educational sciences field between the years of 2005 and 2013. The study sample included a total of 50 articles: 29 scale development and 21 scale adaptation studies. As a result of the study, it was determined that half of the adaptation studies reported data by following the accurate steps. These adaptation studies did not provide information about the decision to adapt the scale, half of them did not mention whether they obtained permission for adaptation, while none of the studies gave information about structural equivalency related to culture, and a majority of them changed the factor structure of the scale and removed too many items. Although the translators had sufficient skills, only three studies employed evaluation and assessment experts. Also, all of these studies conducted reliability analysis. Also, all studies used the Cronbach's alpha method, equivalent splits and test-retest methods. Most of the studies used the EFA. In the adaptation studies, it was not reported that the CFA was used. It was determined that none of these studies conducted a statistical analysis to determine language equivalency.

According to Hambleton et al. (2009), there will be more scale adaptation studies conducted in the future due to the popularization of using the same scales in the international arena, the necessity of tests that fit international assessment and evaluation criteria, and the interest in cross-cultural studies. Scale adaptations are advantageous in financial and temporal grounds. There is no dominantly accepted method in the relevant literature related to scale adaptation studies. This situation improved in 2000s as a result of the developments in measurement and evaluation techniques (SPSS, LISREL, etc.) (Hambleton et al. 2009).

Aim of the Study

With regard to scale adaptation studies in the field of science education in Turkey, only five study done focusing only papers published in short period of time in limited number of journals. Based on our meta-synthesis study, we aimed to provide more conclusive results by reviewing all of the physical science education journals beginning of their first issue. Therefore, aim of this study was to examine scale adaptation studies in Turkey in sciences education area considering content and methodology, and to determine the stages commonly used in the scale adaptation process using the meta-synthesis method. With this objective, this study sought answers for the research questions below:

- 1) What is the situation of scale adaptation studies in Turkey in physical sciences education area in terms of content (area of scale, the year when the scale is published, type of scale, number of items, general reliability coefficient, number of samples and sample level)?
- 2) What is the situation of scale adaptation studies in Turkey in physical sciences education area in terms of methodology?
- 3) What are the frequently followed steps in scale adaptation studies in the field of physical sciences education in Turkey?

Method

This study used the meta-synthesis method which is called thematic content analysis (Walsh and Downe 2005). Calik and Sozbilir (2014) classified the content analysis into three categories: meta-synthesis, meta-analysis, and descriptive content analysis. Thematic content analysis is a useful tool for identifying, analyzing and reporting themes in detail (Braun and Clark 2006). Meta-synthesis is a critical review of different studies conducted on the same subject, and the summary of its results (Calik and Sozbilir 2014; Higgins 2016).

The Population and Sample of This Study

The study population included 145 free journals which were published in the field of physical sciences education in Turkey, and full texts of which are accessible in Istanbul University Library online. The researchers examined these journals from their first issue to April 2016. To access the articles selected for the sample on the websites of the journals, the researcher used these keywords: *scale, scale adaptation, physical sciences, Turkish form, adapting to Turkish, scale adaptation study, adaptation of scale to Turkish, study for adaptation to Turkish, creating a Turkish form, physics, chemistry, biology, physical sciences, science, teacher, pre-service teacher, primary school, secondary school, high school and university students.*

The content analysis found 25 journals containing 46 scale adaptation studies. The journals including the adaptation studies that were analyzed in context of this study were categorized as ULAKBIM (Turkish Academic Network and Information Center), EBSCO, ERIC, SSCI and other indices (Table 1) before examination. In this examination, the indices DOAJ, ACARINDEX, ASOS, OAJI, DRJI, Arastirmax, SIS, IBSS, Web of Science, ERA, Google Scholar and ISI were classified under the title of "other indices". Of the analyzed journals, 81 (46%) are scanned by ULAKBIM, 39 (22%) are scanned by EBSCO, 3 (2%) are scanned by SSCI, 8 (%5) are scanned by ERIC, and 44 (25%) are scanned by other indices.

Table 1. Population index

Indexing	Number of studies
SSCI	3 (%2)
ERIC	8 (%5)
ULAKBIM	81 (%46)
EBSCO	39 (%22)
Other	44 (%25)

*Some journals are scanned by multiple indices.

This study examined a total of 43 articles published in 25 journals which were selected from this population by using the criterion sampling method, one of the purposeful sampling methods (Table 2). Some of these articles include multiple scale adaptations. Thus, the sample of this study included 46 scale adaptation studies (see Appendix). According to Buyukozturk (2009), the units meeting the criteria which were determined by researchers considering the study objective are included in the sample in criterion sampling.

The inclusion criteria for articles:

1. Having been published until April 2016 without any restriction for the initial date,
2. Having been scanned by SSCI, EBSCO, ERIC or other indices in the field of education,
3. Having Turkish origin,
4. Being free of charge,
5. Providing online access to the full manuscript of the articles in the journal,
6. Presence of a scale adaptation study that is conducted in relation with them,
7. Being included in any of physics, chemistry, biology and science teaching areas,
8. Having been conducted with a sample including pre-school, primary school, middle school, high school or university students.

Table 2. Information of journals which comprised the study sample

Journal Name	First Issue (Year)	Last Issue (Year)	Indexing	Number of adaptation studies	
				Year	Number of studies
1 Bilgisayar ve Eğitim Araştırmaları Dergisi	2013	2016	ULAKBIM	2013	1
2 Buca Eğitim Fakültesi Dergisi	1994	2015	EBSCO	2010	1
3 Eğitim Bilimleri ve Uygulama	2002	2014	EBSCO	2002	1
				2003	1
4 Eğitim ve Bilim	1976	2016	ULAKBIM	2005	1
				2013	1
				2014	1
5 Eğitimde Kuram ve Uygulama	2005	2016	EBSCO	2016	1
6 Erzincan Üniversitesi Eğitim Fakültesi	2002	2015	ULAKBIM	2013	2
			EBSCO	2015	1
7 Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi	2000	2015	ULAKBIM	2011	1
8 Gazi Üniversitesi Eğitim Fakültesi Dergisi	1985	2015	ULAKBIM	2013	1
9 Hacettepe Üniversitesi Eğitim Fakültesi Dergisi	1986	2016	ULAKBIM	2004	1
			EBSCO	2007	1
10 International Journal of Assessment Tools in Education	2014	2016	Other	2015	1
11 International Journal of Human Sciences	2004	2016	ULAKBIM	2016	2
			EBSCO		
12 İlköğretim Online	2002	2016	ULAKBIM	2007	1
			EBSCO	2013	1
				2014	2
				2015	2
13 İnönü Üniversitesi Eğitim Fakültesi Dergisi	2007	2016	ULAKBIM	2010	1
			EBSCO		
14 Kırşehir Eğitim Fakültesi Dergisi	2000	2016	ULAKBIM	2009	1
				2013	2
15 Kuram ve Uygulamada Eğitim Bilimleri	2001	2016	ULAKBIM	2004	1
			EBSCO	2005	1
			SSCI	2011	1
			ERIC		
16 Kuram ve Uygulamada Eğitim Yönetimi	1995	2016	ULAKBIM	2008	1
17 Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi	2011	2015	ULAKBIM	2015	1
18 Milli Eğitim Dergisi	1971	2015	ULAKBIM	2002	1
				2013	1
19 Pegem Eğitim ve Öğretim Dergisi	2011	2016	Other	2015	1
20 Trakya Üniversitesi Eğitim Fakültesi Dergisi	2011	2016	ULAKBIM	2013	1
21 Turkish Journal of Education (TURJE)	2012	2016	ULAKBIM	2013	1
22 Turkish Studies	2006	2016	ULAKBIM	2015	1
			EBSCO		
23 Türk Eğitim Bilimleri Dergisi	2003	2014	Other	2009	1
24 Türk Fen Eğitimi Dergisi	2004	2015	ULAKBIM	2010	1
			EBSCO	2011	1
				2012	1
				2013	1
				2014	1
25 Uşak Üniversitesi Sosyal Bilimler Dergisi	2008	2016	ULAKBIM	2012	1
				Total	46

Data Collection Tool

Articles meeting the inclusion criteria were analyzed using three forms which were developed by researchers: "Sample Analysis Table" (Table 3), "Scale Information" (Table 4) and "The Analyses Performed after the Pilot Study" (Table 5). Final versions of these forms were included in the study after the researchers consulted educational measurement experts for their opinions. The creation of these forms is important for the study in

terms of meeting the validity and reliability criteria (Buyukozturk 2009). These forms allowed the researchers to standardize the process and carry out an evaluation according to the same criteria.

Table 3. Sample analysis table

Scale	Sample											
	Elem. sch. (1-4 grades)			Middle school (5-8grade)			High school (9-12 grade)			Pre-service teachers	In-service teachers	Number of sample
	Science	Physics	Chemistry	Science	Physics	Chemistry	Biology	Science	Physics			
Scale-1												
Scale-2												
Scale-3												
.												
.												
.												

Table 4. Scale information

Scales	Scale								Publication year	Scale type			Number of item	Reliability coefficient			
	Likert				Multiple choices					Original scale							
	4	5	7	9	Dichotomic	Original scale	Adapted scale	Adapted scale									
Scale-1																	
Scale-2																	
Scale-3																	
.																	
.																	
.																	

Table 5. The analyses performed after the pilot study

Scale	KMO	Barlett 27% sub-above group	Item discrimination	Reliability		Validity	
				Cronbach Alpha	KR 20/21	Test re-test	Parallel forms
Scale -1							
Scale -2							
Scale -3							
Content	CFA	EFA				0-200	201-400
Criterion	Concurrent					401-600	601-800
	Predictive					801-1000	1001 and above

Data Analysis

The data obtained in this study were analyzed considering the following steps (Sandelowski and Barroso 2007):

1. Designing the synthesis. In this step, the study designs a unique meta-synthesis which will be able to fill a gap in the literature. This meta-synthesis study revealed the purpose and the study question after conducting a literature review on whether there are similar studies.

2. Deciding what is important and necessary for synthesis. In this step, researchers determine the way to collect the studies that are needed for meta-synthesis. This study revealed the study population, keywords and inclusion criteria after the researchers' meticulous assessments.

3. Inclusion of studies. In this step, studies to be included in the meta-synthesis are critically analyzed, and the studies to be included in the meta-synthesis are determined. Of the adaptation studies which were accessed using the chosen keywords in this study, 46 adaptation studies meeting the inclusion criteria were included in the meta-synthesis.

4. Determining the relations of the studies with each other. In this stage, the similarities and differences of the studies included in the meta-synthesis are demonstrated. In this study, the studies included in the meta-synthesis were analyzed using three forms: Sample Analysis Table, Scale Information and the Analyses Performed after Pilot Study.

5. Creating a qualitative meta-synthesis. In this step, an inductive method is followed and the similarities of the studies are combined. This study examined and summarized the data collected with the three forms as data collection tools.

6. Expressing and presenting meta-synthesis. In this step, interpretive explanations of the results, which were achieved as a result of the meta-synthesis, are asserted. This study's findings were associated with the literature, and it was revealed which steps should be taken regarding the scale adaptation process.

To conduct a valid and reliable meta-synthesis study, it is required for the researchers to objectively report the methods followed in the study to the readers, to express inclusion criteria clearly, to include at least 10 studies, selected using the purposeful sampling method, in the meta-synthesis and to examine each of the studies without disrupting the integrity. In this way, the cogency of the study is strengthened (Sandelowski, Docherty and Emden 1997). In this study, the researchers expressed all steps followed in the whole meta-synthesis process in detail, and a total of 46 studies included in the meta-synthesis were objectively analyzed.

Results

This study analyzed articles included in the meta-synthesis method within the scope of the research questions. Initially, content analyses were performed on the articles, which were analyzed for the first study question, by using the "Sample Analysis Table" and "Scale Information" forms. Data obtained are explained below:

Area of Scale

This study assessed adaptation scales in 8 different areas: attitude, self-efficacy, anxiety, motivation, interest, perception, epistemological belief and other areas. In this study, learning environment, competence, distinction between science and pseudoscience, self-regulation, cognitive evaluation, scientific creativity, behavior, learning styles, student assessment in the period of education, argumentation, learning engagement in science, motivation and learning strategies, constructivist thinking, sense of learning and teaching, teacher competence and engagement to integrative STEM teaching areas are discussed as "other areas". As a result of the analysis performed, it was determined that 8 (17%) adaptation studies were in the attitude area, 10 (22%) adaptation studies were in the self-efficacy area, 6 (13%) adaptation studies were in the motivation area, 4 (9%) adaptation studies were in the perception area, 5 (11%) adaptation studies were in the epistemological belief area, 2 (4%) adaptation studies were in the engagement area, and 11 (24%) adaptation studies were in the other areas. No studies in the area of interest have been encountered. Moreover, there is no scale adaptation study about anxiety.

Publishing Year

The analysis which was performed according to publishing years found that there were 1 (2%), 1 (2%), 3 (7%), 2 (4%), 2 (4%), 1 (2%), 2 (4%), 3 (7%), 3 (7%), 2 (4%), 13 (28%), 3 (7%), 7 (15%) and 3 (7%) adaptation studies in 2002, 2003, 2004, 2005, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 and in April 2016, respectively (Figure 1).

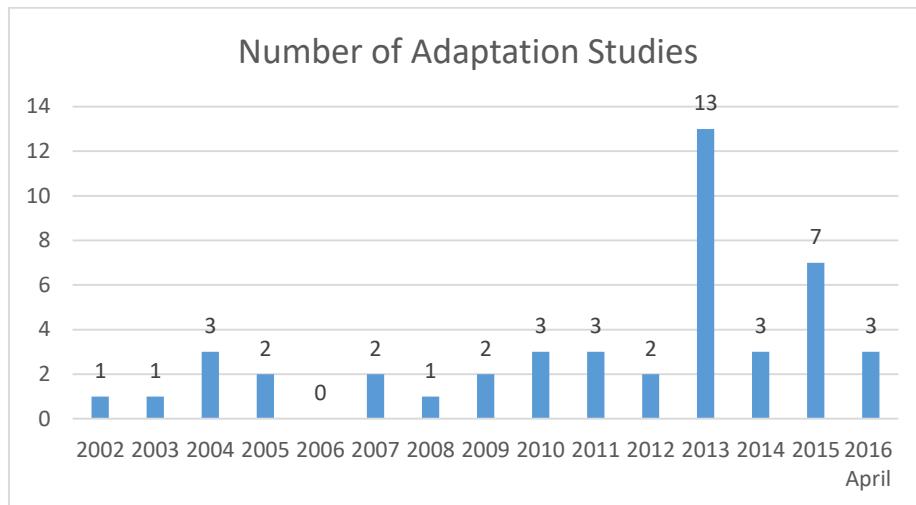


Figure 1. Adaptation studies by year

Type of Scale

Adaptation scales were analyzed under three categories: likert, multiple choice, dichotomous (two-option). Likert-type scales were analyzed under five sub-categories: 4-point, 5-point, 6-point, 7-point, and 9-point. As a result of this analysis, it was found that there were three 4-point likert-type adaptation studies, 29 5-point likert-type adaptation studies, one 6-point likert-type adaptation study and seven 7-point likert-type adaptation studies. There were no adaptation study under the categories of 9-point likert scale and multiple choice. However, a total of four adaptation scales were included in the category of yes-no. The researchers also determined that one study did not mention the type of the scale.

Number of Items

According to number of items, the adaptation scales were examined under two categories: number of items in the original scale and number of items in the adapted version. After this examination, it was found that three studies did not mention number of items in the original scale, and two studies did not mention number of items in the adapted scale. The researchers determined that there was a decrease in number of items in 26 studies.

General Reliability Coefficient

According to general reliability coefficients, the adaptation scales were examined under two categories: reliability coefficient of the original scale and reliability coefficient of the adapted version. According to the analysis results, the researchers did not determine any reliability coefficient regarding the original scale in 39 studies, while 7 studies stated the reliability coefficient of the original scale. Moreover, it was determined that the reliability coefficient of the original scale was not included in 6 studies; therefore, the reliability coefficient of 40 adapted scales was accessed. It was determined that the reliability coefficients of the adapted scales in which the (original version) reliability coefficient took place differed between 0.81 and 0.96. It was also determined that, of the seven studies which gave both the original and the adapted scales' reliability coefficients, as a result of adaptation, five studies experienced a decrease in the reliability coefficient, one experienced an increase in the reliability coefficient, and there was no change in one study. It was found that the reliability coefficients of the adapted scales were calculated between 0.71 and 0.94.

Sample Level and Number of Samples

The sample level was categorized as primary school (1-4th grade), middle school (5-8th grade) and high school (9-12th grade) students, pre-service teachers and in-service teachers. Some studies used more than one study sample (for example, using teachers and teacher candidates together). The researchers did not encounter any adaptation study having a sample at the primary school level. A total of 10 and 8 adaptation studies were determined at the middle and high school levels, respectively. Pre-service teachers were classified to be physical sciences, physics, chemistry and biology. A total of 20, 3, 5, and 5 studies were adapted for teachers in the field of physical sciences, physics, chemistry, and biology, respectively. In-service teachers were selected as a sample in 7 studies. Some parts of the examined articles adapted the scale for more than one sample level (Figure 2).

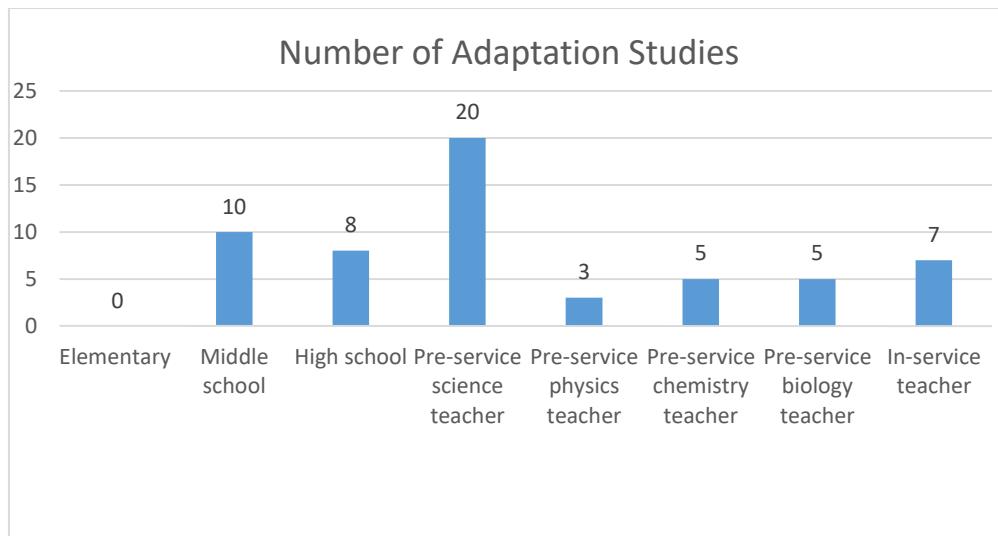


Figure 2. Adaptation studies by sample level

According to sample sizes, adaptation studies were examined under six categories: 0 to 200, 201 to 400, 401 to 600, 601 to 800, 801 to 1000 and 1000 and over. As a result of this examination, it was found that there were 4 (9%), 20 (43%), 11 (24%), 5 (11%), 2 (4%), and 4 (9%) adaptation studies including 0 to 200, 201 to 400, 401 to 600, 601 to 800, 801 to 1000, and 1000 and over respectively. The researchers determined that studies including the number of items of the Turkish form of the scale (44 studies) had seven or more times the sample size for the pilot study (Figure 3).

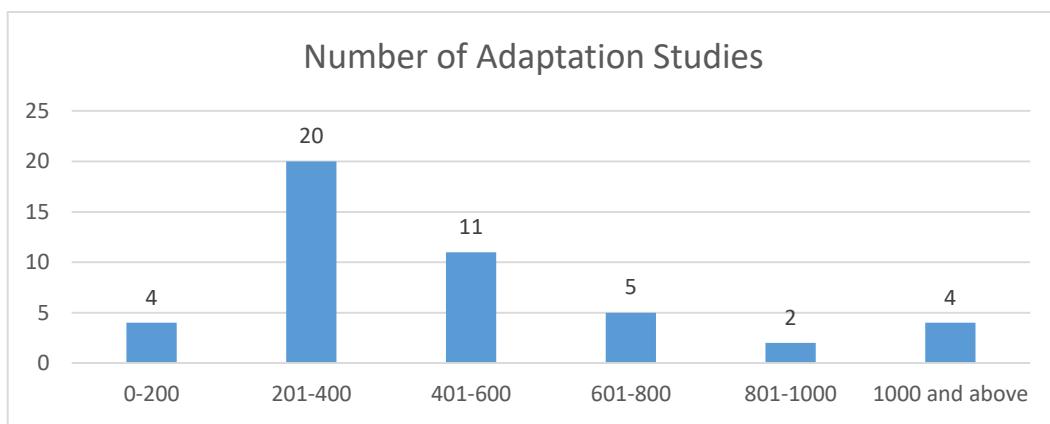


Figure 3. Adaptation studies by sample size

The content analyses were performed on the adaptation articles, which were analyzed within the scope of the second research question, using the “Analyses Performed after the Pilot Study” form. The collected findings are explained below:

For all of the analyzed adaptation studies, the Cronbach's alpha was calculated as an internal consistency coefficient for reliability. Only two (2) studies calculated both the KR-20 and Cronbach's alpha value. The test-retest and split-half reliability methods were calculated in only 15% of studies. Of the 46 studies included in this study, only 28 performed the CFA. The number of studies in which the EFA and CFA analyses were used together was 18. Of the analyzed studies, 5 performed only the EFA and did not perform the CFA. Moreover, 10 studies performed only the CFA and did not perform the EFA. In 8 studies, the validity analysis was not mentioned; it was only stated that the factor analysis was performed (Table 6). It was determined that the KMO and Barlett values were not given in more than 50% of the analyzed studies.

Table 6. Statistical analyses after the pilot study

Statistical Analysis		f
KMO		21
Bartlett		21
Item discrimination	27% sub-above groups	18
	Item correlation	31
Validity	Construct	10
	EFA	5
	CFA+EFA	18
	Unknown	8
	Criterion	2
	Concurrent	2
	Predictive	0
	Unknown	3
	Content	
Reliability	Internal Coefficient	Alpha 46
		KR-20/21 2
	Split-half	7
	Test re-test	8

The adaptation articles that were examined in framework of the third research question were analyzed in three stages (Table 7). Here are the findings derived from these analyses:

1. Preparation for the adaptation study

- Determining the need,
- Selected the scale to be translated,
- Obtaining the permissions required for the adaptation study,
- Forming the translation team.

2. Language validity

- Translating scale items from the original language to Turkish,
- Translating scale instructions from the original language to Turkish,
- Analyzing and comparing the translations,
- Conducting a backward translation,
- Giving its first form to the scale translated to Turkish,
- Implementing the language validity,
- Performing statistical analysis for the language validity,
- Giving its first form to the scale of which the language validity is performed.

3. Pilot study

- Obtaining the permissions required for the pilot study,
- Conducting the pilot study,
- Performing a statistical analysis after the pilot study,
- Attaching the final version of the scale to the article.

Table 7. Analysis of the stages in adaptation studies

Stages		Yes f	No f
Preparation for the adaptation study	Determining the need	46	0
	Selecting the scale to be translated	46	0
	Obtaining the permissions required for the adaptation study	34	12
Language validity	Forming the translation team.	46	0
	Translating scale items from the original language to Turkish	46	0
	Translating scale instructions from the original language to Turkish	0	46
	Analyzing and comparing the translations	46	0
	Conducting a backward translation	32	14
	Giving its first form to the scale translated to Turkish	46	0
	Implementing the language validity	20	26
	Performing statistical analysis for the language validity	19	27
	Giving its first form to the scale of which the language validity is performed	46	0
	Obtaining the permissions required for the pilot study,	0	46
Pilot study	Conducting the pilot study	46	0
	Performing a statistical analysis after the pilot study	46	0
	Attaching the final version of the scale to the article.	24	22

Preparation for the Adaptation Study

This stage was assessed in four aspects. Determining the need, selecting the scale to be translated and forming the translation team were performed for all studies. However, 74% of the studies reported the stage of obtaining the necessary permissions, one of the most important stages to initiate an adaptation study, and 26% did not obtain these permissions.

Language Validity

Translating scale items from the original language to Turkish, performing language consistency (analyzing and comparing translations), giving its first form to the translated test, and giving its first form to the scale of which the language validity is performed were available in all studies. Of the studies, 70% did a backward translation. In terms of language validity, 43% of the analyzed studies were carried out in philology or English language teaching departments that have a command of both languages. Of the adaptation studies which were analyzed, 41% performed a statistical analysis for the language validity. No studies mentioned the translation of the instructions.

Pilot Study

No studies mentioned obtaining permission for the pilot study. All studies conducted a pilot study and performed statistical analyses for it. Only 52% of studies included the original scale.

Discussion and Conclusion

In this study, the analyzed adaptation studies were examined in three stages. Firstly, "scale information" were revealed in all of the analyzed studies. Areas of the scales, publishing dates, types, number of items included and general reliability coefficients were determined. Then, *sample analyses* were performed on the analyzed articles. In this way, sample size and sample type were determined. Finally, the method followed by each study was analyzed after the pilot study.

This study found that adaptation studies are mostly conducted in the self-efficacy area. As it was reported in the studies of Karaca et al. (2008), Boztunc - Ozturk, Eroglu and Kelecioglu (2015), Delice and Ergene (2015) and Gul and Sozbilir (2015b), this study found that the internal consistency coefficients were calculated in all examined studies. It was seen that the likert type was used significantly more than the dicothomic scales. While the Cronbach's alpha value was calculated for reliability in the likert-type scales, the dichotomous scales calculated KR-20/21 for reliability (Secer 2015). In four studies determined as dichotomous, two calculated the KR-20 and Cronbach's alpha together although the KR-20 value should be calculated as a reliability coefficient. However, the other two studies only calculated the Cronbach's alpha value, but did not calculate the KR-20 value. This showed that some researchers did not have sufficient statistical information on internal consistency calculations. This study determined that there were no significant differences between the internal consistency coefficients of the original scale and the adapted scale. It was found that some items were excluded from the scale as a result of the adaptation study, while no items were added to the scales in the adaptation studies. These findings are mainly consistent with the studies of Boztunc-Ozturk et al. (2015).

This study showed that pre-service physical sciences teachers were selected as a sample level in the adaptation studies, and no sample at the primary school level was used. Moreover, it was found that the sample size was at least 7 times or bigger than the number of items. It was determined that the researchers exercised due diligence regarding this issue. However, Delice and Ergene (2015) reported that some studies which they analyzed had a deficient sample size while it is supposed to be five times bigger than the number of items.

According to the analysis performed by the publishing date, it was found that the first adaptation study was conducted in the field of physical sciences in 2012. In Turkey, teacher training programs were reformed as a 4-year program by the Council of Higher Education (YOK) in 1989. Gazi and Ege Universities' Education Faculties began to train physical sciences teachers in 1992. Between the years of 1992 and 1998, these faculties trained physical science teachers according to the curricula that they prepared. In the 1998 – 1999 academic year, courses and their content followed in the faculties of education were reviewed, and the prepared physical science teaching curricula was sent to other faculties (Kizilcaoglu 2006; Meric and Tezcan 2005). Therefore, scale adaptation and development studies in the field of physical sciences education began to be conducted after the 2000s. While only 2 adaptations of articles were conducted before 2004, there were no adaptation scales in 2006. In 2013 and 2015, the number of adaptations of article studies increased dramatically.

As a result of the analysis, which was performed to determine the stages followed in adaptation studies, this study found that these studies gathered under three main headings of "preparation for the adaptation study", "language validity" and "pilot study". The "preparation for the adaptation study" stage includes determining the needs, selecting the scale to be translated, obtaining the necessary permissions and forming the translation team. This study determined that all analyzed studies had steps of determining the needs, selecting the scale and forming the translation team. However, some studies did not obtain permission from the developers of the original scale. Similarly, Cum and Koc (2013) reported in their studies that half of the researchers did not obtain permission for adaptation of the scale.

The literature has stated that the most important issue is "validity of language" while adapting the scale (Delice and Ergene 2015; Tsagari and Floros 2013). Backward translation, the most important step in determining the validity of the language, and the examination of an expert team who makes this translation have been frequently emphasized in the literature (Cha et al. 2007; Grisay 2000; Maxwell 1996; Tsagari and Floros 2013). This study determined that the analyzed adaptation studies did not mention this stage sufficiently, it was discussed superficially, and even 14 adaptation studies (30% of the adaptation studies) did not give detailed information about the backward translation. The fact that the pilot study stage, which is required to be performed on a group having a command of both two languages and cultures for the validity of the language, was determined in only 20 of the adaptation studies (43% of the analyzed studies) is an indicator that in Turkey, researchers do not give due importance and sensitivity to the validity of the language. Savasir (1994) reported that less than half of the 33 scale adaptation studies published in the Journal of Psychology and the Turkish Journal of Psychiatry did not give information about the translation; other studies mentioned this issue only briefly. Cum and Koc (2013),

Delice and Ergene (2015), Guvendir and Ozkan (2015) examined adaptation studies and found that the studies did not form an expert group having a command of both languages and cultures for the validity of the language. Being largely in agreement with the study of Guvendir and Ozkan (2015), the present study determined that no researcher adapted the instructions to the target language.

The last stage of the adaptation studies, the “pilot study”, was available in all analyzed studies. This study concluded that all studies did not obtain relevant official permissions for the pilot study. Similarly, Delice and Ergene (2015) reported that more than half of the adaptation studies did not obtain necessary permissions. While it is commonly preferred by the researchers to perform only the CFA analyses in scale adaptation studies, there are studies in the literature where both analysis methods were preferred (Atilgan 2015). At the end of this study, being largely in agreement with the studies of Boztunc-Ozturk et al. (2015), it was determined that researchers preferred to use the EFA and CFA methods together. Guvendir and Ozkan (2015), Cum and Koc (2013) found in their studies that the researchers preferred the EFA method to the CFA method.

A study of Boztunc-Ozturk et al. (2014) determined that adaptation studies generally did not examine the scale validity. Similarly, the present study also found that the researchers calculated the construct validity, but did not perform the analyses required for the scale validity. It was observed that more than half of the studies gave the adapted scale together with the article.

Similarly; Tavsancil, Guler and Ayan (2014) examined adaptation studies conducted in the field of education in Turkey. They determined that most of the followed steps had wrong and incomplete information. Cum and Koc (2013) also examined adaptation studies and reported that the followed steps in these studies were inappropriate. Therefore, researchers who will conduct an adaptation study should have basic statistical skills, and studies should be conducted according to test development stages which are nationally/internationally determined. These will help future adaptation studies be more productive.

Limitations of this study include focusing on only scale adaptation studies in the field of science education published in the journals and excluding Master and PhD Theses. Especially in Turkey, becoming very popular to develop or adapt a scale for Master and PhD theses. Therefore, future studies could be focus on detailed examinations of these theses in terms of scale adaptation and scale development studies in the field of science, mathematics and technology education.

References

- Akbas, G., & Korkmaz, L. (2007). Ölçek uyarlaması [Scale adaptation]. *Türk Psikoloji Bülteni*, 13(40), 15-16.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Atilgan, H. (2015). *Eğitimde ölçme ve değerlendirme* [Assessment in education]. Ankara: Anı Yayıncılık.
- Boztunc Ozturk, N., Eroglu, M.G., & Kelecioglu, H. (2015). Eğitim alanında yapılan ölçek uyarlama makalelerinin incelenmesi [Examination of scale adaptation articles in the field of education]. *Eğitim ve Bilim*, 40(178), 123-137.
- Braun V., & Clarke V. (2006). Using thematic analysis in psychology. *Qualitative Research Psychology*, 2006(3), 77–101.
- Buyukozturk, Ş. (2009). *Sosyal bilimler için veri analizi el kitabı* [Data analysis handbook for social sciences] (10th ed.). Ankara: Pegem Akademi.
- Calik, M., & Sozbilir, M. (2014). İçerik analizin parametreleri [Parameters for content analysis]. *Eğitim ve Bilim*, 39(174), 33-38.
- Cha, E. S., Kim, K. H., & Erlen, J. A. (2007). Translation of scales in cross-cultural research: Issues and techniques. *Journal of Advanced Nursing*, 58(4), 386–395.
- Costello, A.B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 10(7), 1-9.
- Cum, S., & Koc, N. (2013). Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi [An examination of scale development and adaptation studies published in psychology and educational sciences journals in Turkey]. *Eğitim Bilimleri ve Uygulama*, 12 (24), 115-135.
- de Vijver, F. V., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1(2), 89-99.

- Delice, A., & Ergene, Ö. (2015). Ölçek geliştirme ve uyarlama çalışmalarının incelenmesi: Matematik eğitimi makaleleri örneği [Examining scale development and adaptation studies: An example of mathematics education articles]. *Karaelmas Journal of Educational Sciences*, 3, 60-75.
- Deniz, K. Z. (2007). The adaptation of psychological scales. *Ankara University, Journal of Faculty of Educational sciences*, 40(1), 1-16.
- Ebbs, D., & Korsnakova, P. (2015). *Translation and translation verification for TIMSS 2015*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6(4), 304.
- Gomleksiz, M., & Erkan, S. (2010). *Eğitimde ölçme ve değerlendirme* [Assessment in education] (2nd ed.). Ankara: Nobel Yayın Dağıtım.
- Grisay, A. (2000). Translation and cultural appropriateness of the test and survey material. *PISA 2000 Technical Report*, 57-70.
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, 61, 29-45.
- Gul, Ş., & Sozbilir, M. (2015a). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi [Thematic content analysis for scale development researches in science and mathematics education]. *Eğitim ve Bilim*, 40(178), 85-102.
- Gul, Ş., & Sozbilir, M. (2015b). Biology education research trends in Turkey. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(1), 93-109.
- Guvendir, M. A., & Ozkan, Y. Ö. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi [Examining the articles on scale development and adaptation in the scientific journals published in the field of education in Turkey]. *Elektronik Sosyal Bilimler Dergisi*, 14(52), 23-33.
- Hambleton, R.K., Merenda, P. F., & Speilberger, C.D. (2009). *Adapting educational and psychological tests for cross-cultural assessment*. London: Lawrence Erlbaum Publishers.
- Higgins, S. (2016). Meta-synthesis and comparative meta-analysis of education research findings: Some risks and benefits. *Review of Education*, 4(1), 31-53.
- Karaca, E., Yurdabakan, I., Cetin, B., Nartgun, Z., Bicak, B., & Gomleksiz, M. (2008). *Eğitimde ölçme ve değerlendirme* [Assessment in education]. Ankara: Nobel Akademik Yayıncılık.
- Kızılcaoglu, A. (2006). Eğitim Fakültelerinde yeniden yapılandırma sürecine ilişkin eleştiriler ve öneriler [Criticisms and suggestions about the restructuring process in education faculties]. *Balıkesir Üniversitesi, Sosyal Bilimler Enstitüsü Dergisi*, 8(14), 132-140.
- Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.
- Meric, G., & Tezcan, R. (2005). Fen bilgisi öğretmeni yetiştirme programlarının örnek ülkeler kapsamında değerlendirilmesi [Evaluation of science education training programs within sample countries]. *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 7(1), 62-82.
- OECD (2017). The Programme for International Student Assessment. Retrieved from <http://www.oecd.org/pisa/aboutpisa/>
- Plake, B. S., & Wise, L.L. (2014). What is the role and importance of the revised AERA, APA, NCME standards for educational and psychological testing? *Educational Measurement: Issues and Practice*, 33(4), 4-12.
- Polat, M., Gonen, E., Parlak, B., Yildirim, A., & Ozgurluk, B. (2016). *TIMSS 2015 ulusal matematik ve fen bilimleri ön raporu 4. ve 8. sınıflar*. [TIMSS 2015 national mathematics and science preliminary report for grades 4 and 8]. Ankara: Milli Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü.
- Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. New York: Springer.
- Sandelowski, M., Docherty, S., & Emden, C. (1997). Focus on qualitative methods qualitative metasynthesis: issues and techniques. *Research in Nursing and Health*, 20, 365-372.
- Savasir, I. (1994). Ölçek uyarlamasındaki sorunlar ve bazı çözüm yolları [Problems of scale adaptation and some solutions]. *Türk Psikoloji Dergisi*, 9(33), 27-32.
- Secer, I. (2015). *Psikolojik Test Geliştirme ve Uyarlama Süreci; SPSS ve LISREL Uygulamaları* [Psychological Test Development and Adaptation Process; SPSS and LISREL Applications]. Ankara: Anı Yayıncılık.
- Seker, H., & Gencdogan, B. (2006). *Psikolojide ve eğitimde ölçme aracı geliştirme* [Scale development in Psychology and education]. Ankara: Nobel Yayın Dağıtım.
- Sireci, S. G. (1995). Using cluster analysis to solve the problem of standard setting. Paper presented at the meeting of the American Psychological Association, New York.

- Tavsancıl, E., Guler, G., & Ayan, C. (2014). *2002-2012 yılları arasında Türkiye'de geliştirilen bazı tutum ölçeği geliştirme çalışmalarıının ölçek geliştirme süreci açısından incelenmesi* [Investigation of some attitude scale development works developed in Turkey between 2002-2012]. IV. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı), Hacettepe Üniversitesi, Ankara.
- Tsagari, D., & Floros, G. (2013). *Translation in language teaching and assessment*. Cambridge: Cambridge Scholars Publishing.
- Turgut, M.F., & Baykul Y. (2012). *Eğitimde ölçme ve değerlendirme* [Assessment in education]. Ankara: Pegem Yayıncılık.
- Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: A literature review. *Journal of Advanced Nursing*, 50(2), 204-211.

Author Information

Behiye Akçay

Istanbul University-Cerrahpasa
İstanbul, Turkey
e-mail: bbezir@gmail.com

Aydın Tiryaki

Istanbul University-Cerrahpasa
İstanbul, Turkey

Burcu Gelen

Istanbul University-Cerrahpasa
İstanbul, Turkey

İbrahim Benek

Istanbul University-Cerrahpasa
İstanbul, Turkey

Appendix. List of Adaptation Studies

Scale-1	Yılmaz, M., Gürçay, D., & Ekici, G. (2007). Akademik Özyeterlik Ölçeğinin Türkçe'ye Uyarlanması. <i>Hacettepe Üniversitesi Eğitim Fakültesi Dergisi</i> , 33, 253-259.
Scale -2	Yılmaz, M., Köseoglu, P., Gerçek, C., & Soran, H. (2004). Yabancı Dilde Hazırlanan Bir Öğretmen Öz-Yeterlik Ölçeğinin Türkçeye Uygulanması. <i>Hacettepe Üniversitesi Eğitim Fakültesi Dergisi</i> , 27, 260-267.
Scale -3	Yapalak, S., & Ilgaz, G. (2013). The Adaptation Of “Attitudes Toward Research (ATR)” Scale Into Turkish. <i>Journal Of Kırşehir Education Faculty</i> , 14(2), 79-90.
Ölçek-4	Tosun, C., Şenocak, E., & Taşkesenligil, Y. (2009). Endüstriyel Kimya Dersi Öğrenme Ortamı Ölçeğinin Türkçeye Uyarlanması Çalışması. <i>Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi</i> , 10(2), 111-117.
Scale -5	Özel, M., Timur, B., Timur, S., & Bilen, K. (2013). Öğretim Elemanlarının Pedagojik Alan Bilgilerini Değerlendirme Anketinin Türkçeye Uyarlanması Çalışması. <i>Journal Of Kırşehir Education Faculty</i> , 14(1), 407-428.
Scale -6	Dindar, A. Ç., & Geban, Ö. (2015). Fen Bilimleri Motivasyon Ölçeğinin Türkçe'ye ve Kimya'ya Uyarlanması: Geçerlilik Çalışması. <i>Pegem Eğitim ve Öğretim Dergisi</i> , 5(1), 15-34.
Scale -7	Çetinkaya, E. K., Şimşek, C. L., & Çalışkan, H. (2013). Bilim ve Sözde-Bilim Ayrımı İçin Bir Ölçek Uyarlama Çalışması. <i>Trakya Üniversitesi Eğitim Fakültesi Dergisi</i> , 3(2), 31-43.
Scale -8	Aydın, S., Keskin, M. Ö., & Yel, M. (2015). Öz-Düzenleme Ölçeğinin Türkçe Uyarlaması: Geçerlik ve Güvenirlilik Çalışması. <i>Turkish Journal Of Education</i> , 3(1), 24-33.
Scale -9	Terzi, Ş. (2009). Bilişsel Değerlendirme Ölçeği'nin Uyarlanması: Geçerlik ve Güvenirlilik Çalışmaları. <i>Türk Eğitim Bilimleri Dergisi</i> , 7(1), 127-141.
Scale -10	Çeliker, H. D., & Balım, A. G. (2012). Bilimsel Yaratıcılık Ölçeğinin Türkçeye Uyarlama Süreci ve Değerlendirme Ölçütleri. <i>Uşak Üniversitesi Sosyal Bilimler Dergisi</i> , 5(2), 1-21.
Scale -11	Deryakulu, D., & Bıkmaç, H. F. (2003). Bilimsel Epistemolojik İnançlar Ölçeğinin Geçerlik ve Güvenirlilik Çalışması. <i>Eğitim Bilimleri ve Uygulama</i> , 2(4), 243-257.
Scale -12	Timur, S., & Yılmaz, M. (2013). Çevre Davranış Ölçeğinin Türkçe'ye Uyarlanması. <i>Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi</i> , 33(2), 317-333.
Scale -13	Totan, T., İkiz, E., & Karaca, R. (2010). Duygusal Öz-Yeterlik Ölçeğinin Türkçeye Uyaranarak Tek ve Dört Faktörlü Yapısının Psikometrik Özelliklerinin İncelenmesi. <i>Buca Eğitim Fakültesi Dergisi</i> , 28, 71-95.
Scale -14	Candan, D. G., & Gencel, İ. E. (2015). Öğretme Motivasyonu Ölçeği'nin Türkçe'ye Uyarlama Çalışması. <i>Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi</i> , 1(36), 72-89.
Scale -15	Yapıcı, İ. Ü., & Hevedanlı, M. (2013). Biyoloji Öğretiminde BİT Kullanımı Tutum Ölçeğinin Türkçeye Uyarlanması: Geçerlik ve Güvenirlilik Çalışması. <i>Journal Of Computer And Education Research</i> , 1(2), 21-37.
Scale -16	Aypay, A. (2010). General Self-Efficacy (GSE) Scale To Turkish. <i>Inonu University Journal Of The Faculty Of Education</i> , 11(2), 113-131.
Scale -17	Tosun, C. (2013). Kimya Motivasyon Ölçeği-II'nin Türkçe'ye Uyarlanması: Geçerlik ve Güvenirlilik Çalışması. <i>Erzincan Üniversitesi Eğitim Fakültesi Dergisi</i> , 15(1), 173-202.
Scale -18	Tosun, C. (2013). Kimya Motivasyon Ölçeği-II'nin Türkçe'ye Uyarlanması: Geçerlik ve Güvenirlilik Çalışması. <i>Erzincan Üniversitesi Eğitim Fakültesi Dergisi</i> , 15(1), 173-202.
Scale -19	Tosun, C., & Şekerci, A. R. (2015). Sekizinci Sınıf Öğrencilerinin Fen Derslerine Karşı Öz Düzenleme Becerilerinde Motivasyonun Rolü. <i>Erzincan Üniversitesi Eğitim Fakültesi Dergisi</i> , 17(1), 1-29.
Scale -20	Arhan, S., Gültekin, İ. (2013). Türkçe Dersi İlköğretim Programlarında Benimsenen Tematik Yaklaşımın Metin Seçimine Etkileri Yönünden Değerlendirilmesi. <i>Milli Eğitim Dergisi</i> , 119.
Scale -21	Aypay, A. (2011). Epistemolojik İnançlar Ölçeğinin Türkiye Uyarlaması ve Öğretmen Adaylarının Epistemolojik İnançlarının İncelenmesi. <i>Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi</i> , 12(1), 1-15.
Scale -22	Hazır-Bıkmaç, F. (2004). Sınıf Öğretmenlerinin Fen Öğretiminde Öz Yeterlilik İnanç Ölçeğinin Geçerlik ve Güvenirlilik Çalışması. <i>Milli Eğitim Dergisi</i> , 161.
Scale-23	Bıkmaç, F. H. (2002). Fen Öğretiminde Öz-Yeterlik İnanç Ölçeği. <i>Eğitim Bilimleri ve Uygulama</i> , 1(2), 197-210.
Scale -24	Çapa, Y., Çakiroğlu, J., & Sarıkaya, H. (2005). Öğretmen Özyeterlik Ölçeği Türkçe Uyarlamasının Geçerlik ve Güvenirlilik Çalışması. <i>Eğitim ve Bilim</i> , 137(30), 74-81.
Scale -25	Özgüngör, S. (2013). Eğitim Süreci Öğrenci Değerlendirmeleri Ölçeği'nin Türkçeye Uyarlama

	Çalışması ve Etkili Öğretim Elemanlarını Tanımlayan Alt Boyutların Belirlenmesi. <i>Eğitim ve Bilim</i> , 170(38), 79-94.
Scale -26	Buldur, S., & Doğan, A. (2014). Fen ve Teknoloji Dersinde Öğrencilerin Sınıf-İçi Değerlendirme Ortamına İlişkin Algıları Ölçeğinin Türkçeye Uyarlanması. <i>Eğitim ve Bilim</i> , 176(39), 199-211.
Scale -27	Yılmaz, H., & Çavaş, P. H. (2007). Reliability and Validity Study Of The Students' Motivation Toward Science Learning (SMTSL) Questionnaire. <i>Elementary Education Online</i> , 6(3), 430-440.
Scale -28	Çapa-Aydın, Y., Uzuntiryaki-Kondakçı, E., Temli, Y., & Tarkin, A. (2013). Adaptation Of Sources Of Self-Efficacy Inventory Into Turkish. <i>Elementary Education Online</i> , 12(3), 749-758.
Scale -29	Kaya, E., Cetin, P. S., & Erduran, S. (2014). Adaptation Of Two Argumentation Tests Into Turkish. <i>Elementary Education Online</i> , 13(3), 1014-1032.
Scale -30	Kaya, E., Cetin, P. S., & Erduran, S. (2014). Adaptation Of Two Argumentation Tests Into Turkish. <i>Elementary Education Online</i> , 13(3), 1014-1032.
Scale -31	Tosun, C., & Genç, M. (2015). Adaptation Of Science Attitude Scale Developed For Elementary School Students To Turkish: Validity and Reliability Study. <i>Elementary Education Online</i> , 14(3), 946-960.
Scale -32	Yetişir, M. İ., & Ceylan, E. (2015). The Adaptation Of Students' Adaptive Learning Engagement İn Science Scale Into Turkish. <i>Elementary Education Online</i> , 14(2), 657-670.
Scale -33	Bayar, A., & Karamustafaoglu, O. (2015). The Colorado Learning Attitudes About Science Survey (CLASS): The Study Of Validity and Reliability. <i>International Journal Of Assessment Tools İn Education (IJATE)</i> , 2(1), 40-57.
Scale -34	Büyüköztürk, Ş., Akgün, Ö. E., Kahveci, Ö., & Demirel, F. (2004). Gündülenme ve Öğrenme Stratejileri Ölçeğinin Türkçe Formunun Geçerlik ve Güvenirlilik Çalışması. <i>Kuram ve Uygulamada Eğitim Bilimleri</i> , 4(2), 207-239.
Scale -35	Tosun, Ü., & Karadağ, E. (2008). Yapılandırmacı Düşünme Envanterinin Türkçeye Uyarlanması Dil Geçerliği ve Psikometrik İncelemesi. <i>Kuram ve Uygulamada Eğitim Bilimleri</i> , 8(1), 225-264.
Scale -36	Aypay, A. (2011). Öğretme ve Öğrenme Anlayışları Ölçeği'nin Türkiye Uyarlaması ve Epistemolojik İnançlar İle Öğretme ve Öğrenme Anlayışları Arasındaki İlişkiler. <i>Kuram ve Uygulamada Eğitim Bilimleri</i> , 11(1), 7-29.
Scale -37	Baloğlu, N., & Karadağ, E. (2008). Öğretmen Yetkinliğinin Tarihsel Gelişimi ve Ohio Öğretmen Yetkinlik Ölçeği: Türk Kültürüne Uyarlama, Dil Geçerliği ve Faktör Yapısının İncelenmesi. <i>Kuram ve Uygulamada Eğitim Yönetimi</i> , 56, 571-606.
Scale -38	Acat, M. B., Tüken, G., & Karadağ, E. (2010). Bilimsel Epistemolojik İnançlar Ölçeği: Türk Kültürüne Uyarlama, Dil Geçerliği ve Faktör Yapısının İncelenmesi. <i>Türk Fen Eğitimi Dergisi</i> , 7(4), 67-89.
Scale -39	Şenocak, E. (2011). Kimya Dersi Tutum Ölçeğinin Türkçeye Uyarlanması Çalışması. <i>Türk Fen Eğitimi Dergisi</i> , 8(2), 114-129.
Scale -40	Bilici, S. C., Armağan, F. Ö., Çakır, N. K., & Yürük, N. (2012). Astronomi Tutum Ölçeğinin Türkçe'ye Uyarlanması: Geçerlik ve Güvenirlilik Çalışması. <i>Journal Of Turkish Science Education</i> , 9(2), 116-127.
Scale -41	Timur, B., & Taşar, M. F. (2013). Fen Öğretiminde Bilgisayar Kullanımına Yönelik Öz Yeterlik İnancı Ölçeğinin Türkçe'ye Uyarlanması. <i>Türk Fen Eğitimi Dergisi</i> , 10(3), 59-72.
Scale -42	Yaman, S., & Köksal, M. S. (2014). Fen Öğrenmede Zihinsel Risk Alma ve Yordayıcılarına İlişkin Algı Ölçeği Türkçe Formunun Uyarlanması: Geçerlik ve Güvenirlilik Çalışması. <i>Journal Of Turkish Science Education</i> , 11(3), 119-142.
Scale -43	Yıldırım, B., & Selvi, M. (2015). Adaptation Of STEM Attitude Scale To Turkish. <i>Electronic Turkish Studies</i> , 10(3), 1117-1130.
Scale -44	Gülhan, F., & Şahin, F. (2016). Fen-teknoloji-mühendislik-matematik entegrasyonunun (STEM) 5. Sınıf öğrencilerinin bu alanlarla ilgili algı ve tutumlarına etkisi. <i>International Journal Of Human Science</i> , 13(1), 602-620.
Scale -45	Gülhan, F., & Şahin, F. (2016). Fen-teknoloji-mühendislik-matematik entegrasyonunun (STEM) 5. Sınıf öğrencilerinin bu alanlarla ilgili algı ve tutumlarına etkisi. <i>International Journal Of Human Science</i> , 13(1), 602-620.
Scale -46	Hacıömeroğlu, G., & Bulut, A. S. (2016). Integrative STEM Teaching Intention Questionnaire: A Validity and Reliability Study Of The Turkish Form. <i>Eğitimde Kuram ve Uygulama</i> , 12(3), 654-669.