Education and Science

Vol 44 (2019) No 198 361-388

The Use of Exploratory and Confirmatory Factor Analyses: A Document Analysis *

İlhan Koyuncu [1], Abdullah Faruk Kılıç [2]

Abstract

This paper aims to review the scale development research published in Turkey between 2006 and 2016 with regard to their processes of exploratory (EFA) and confirmatory factor analysis (CFA). Within this scope, the distribution of the studies according to years and factor analyses, the extent to which their hypotheses came through for both analysis methods, and the distribution of EFA and CFA results according to the published papers were investigated. In this way, we aimed to infer significant results on the practical application of EFA and CFA which are frequently referred in theory. Hence, the present research incorporates a descriptive document analysis on 131 scale development studies published in the fields of education and social sciences in the journals indexed in The Scientific and Technological Research Council of Turkey (TÜBİTAK) TR Dizin data base between 2006 and 2016. Frequency tables, column charts, histogram, line charts, and measures of central tendency were used to analyse the data. More than 70% of 131 scale development studies were carried out after 2013. The maximum number of studies was published in 2015, and the minimum number in 2008. Only EFA was used in 59 articles, only CFA was used in 4 articles, and both analysis methods were employed in 58 articles. The remaining 10 papers used none of the analyses and included only item analyses. The average sample sizes were 395 participants for EFA and 529 participants for CFA. Sample sizes were accordingly adequate in terms of the factor analyses. EFA and CFA were conducted over the same sample in 36 papers, whereas the analyses were based on different samples in 22 ones. In terms of estimation, 65% of 117 EFA using papers employed Principle Component Analysis (PCA), 2% used Maximum Likelihood Estimation (MLE), 1% used Parallel Analysis (PA), and 3% used Principle Axis Factoring (PAF), while 29% did not specify their method of estimation. The hypotheses of EFA analyses were examined in 93% of the papers. Scree plots were taken into consideration when determing the number of factors in

Keywords

Exploratory Factor Analysis (EFA)
Confirmatory Factor Analysis (CFA)
Document analysis
Scale development
Validity
Construct validity

Article Info

Received: 12.15.2017
Accepted: 02.12.2019
Online Published: 04.22.2019

DOI: 10.15390/EB.2019.7665

* This article is the extended version of the paper titled "Türkiye'de 2006-2016 Yılları Arasında Yayınlanan Ölçek Geliştirme Çalışmalarının Faktör Analizi Açısından İncelenmesi" and presented at the IVth International Eurasian Educational Research Congress.

[1] Adıyaman University, Faculty of Education, Educational Measurement and Evaluation, Turkey, ilhankync@gmail.com
[2] Hacettepe University, Institute of Educational Sciences, Educational Measurement and Evaluation, Turkey, abdullahfarukkilic@gmail.com

50% of the papers. Varimax rotation method was used in 67% of the EFA using papers. The total variance explained was 53.05%. The analytical hypotheses were not examined in 87% of the 62 studies conducted with CFA; 51 of them included track diagrams and 52 included factor loadings. 90% of the papers did not give any information on parameter estimation method. All indices, except the Parsimony Goodness of Fit Index (PGFI), were acceptable as model fit indices in the papers. The results from the papers examined were evaluated in the light of the relevant literature in the last section of the present paper, and some suggestions were made in reference to both theory and practice.

Introduction

Scale development is applied as a systematic process so as to obtain a range of numerical measures by using the indicators of a certain structure in psychology and behavioural sciences. Davidshofer and Murphy (2005) divide scale development process into three parts, namely constructing, standardising, and revising/updating the instrument. While some researchers limit scale development to scale construction, others consider the standardisation within scale development process. According to Crocker and Algina (1986), scale development process includes the stages of identifying the primary purpose for which the test scores will be used, identifying the behaviours that represent the construct or define the domain, preparing a set of test specification, constructing an initial pool of items, pilot test, field test, item analyses, reliability and validity studies, as well as standardisation. Clark and Watson (1995) define the scale development as a process encompassing a clear description of the target structure, a careful construction of the item pool, testing those items on a representative sample, analysing the factors and correlation between items, validity of dimensionality and discrimination. In addition to the steps put forward by Crocker and Algina (1986), the scale development process proposed by Erkuş (2012) includes selecting the technique of scale development, writing out explanations and guidelines, and applying re-test if necessary. Although the definitions and steps differ from each other, the common primary aim of scale development is to construct a valid measure of the underlying psychological construct (Clark & Watson, 1995).

The validity of an assessment instrument involves the process of gathering evidence to support what is inferred from the test scores (Cronbach, 1984). The generally accepted classification by American Psychological Association [APA], American Educational Research Association [AERA] and National Counseling on Measurement in Education [NCME] (1974) grounds the validity evidence on content, construct, and criterion. Content validity is related to whether items constitute a sample of subject and behavioural area (Cronbach & Meehl, 1955). Expert opinion is generally taken to provide evidence for content analysis. Criterion validity is about whether a test measures what is intended to measure (Cureton, 1951). In order to get criterion-based evidence, researchers observe to what extent the scale fits to the other scales in the literature that measure the same construct or the sub-constructs and proved to be reliable and valid. Construct validity shows whether test scores can measure a construct only via the indicators of that construct (Kline, 2016). In other words, construct validity indicates the extent to which a construct is measured by a test or scale. According to Cronbach and Meehl (1955), studies on group differences, correlation matrices, factor analyses, internal construct, conditional changes, and processes can be used to get evidence for construct validity.

Factor analysis is the most widely used method in psychometric assessment (for construct validity) of tests in applied research (Brown, 2015). Factor analysis, according to Kline (2016), is a combination of statistical techniques aiming at reducing or simplifying the data, based on a correlation or covariance matrix. Therefore, compared to simple item analyses, factor analysis is a highly effective way to specify items or factors best explaining a construct (regarding the extent of variance). Two techniques are used in this method of analysis whose foundations were laid in the 1900s (e.g. Spearman,

1904). One of them is exploratory factor analysis and the other is confirmatory factor analysis. Convenience of data type and quality, knowing the distribution of variables and reaching a sufficient sample size are among the important assumptions in factor analysis techniques (Floyd & Widaman, 1995). In exploratory factor analysis, researchers aim to determine the optimal number of factors and to reveal whether there are reasonable indicators for various potential aspects of measured variables/items (Brown, 2015). Confirmatory factor analysis (Jöreskog, 1967, 1971) tests whether a hypothetical factor structure fits to the covariance structure observed for the variables measured (Floyd & Widaman, 1995). Although they require a series of similar assumptions, exploratory and confirmatory factor analyses are multivariate statistical methods with distinctive processes. As the present paper deals with the uses of exploratory and confirmatory factor analyses, the processes and the contents of those processes for each method will be clarified. Thus, the paper will provide a detailed information about exploratory and confirmatory analyses.

The principal procedure of *exploratory factor analysis* (EFA) includes adequacy of sample size, choosing the factor extraction method, examining the Kaiser-Meyer-Olkin (KMO) and Bartlett tests, as well as scree plots and factor loads, choosing the factor rotation method, naming the factors, and reporting the percentage of total variance explained. The recommended sample size for EFA is minimum 300, with observations between 5 and 10 for each variable/item (Comrey & Lee, 1992; Tabachnick & Fidell, 2012). 100 is poor, 200 is fair, 300 is good, 500 is very good, 1000 or more is an excellent sample size according to Comrey and Lee (1992). Even though 10 observations for each variable or 30 observations at least for each factor are recommended as well, greater sample size reduces the error rate in resultant data and makes EFA function much better (Yong & Pearce, 2013). However, Guadagnoli and Velicer (1988) assert that a sample of 150 or more participants can be adequate if factor loads of several items exceed 0.80. In EFA, besides the adequate sample size, it is necessary to find out whether the data fit to the factor analysis. Therefore, KMO and Bartlett tests should be analysed; the KMO value is expected to be greater than 0.50 and Bartlett test should be significant (Hair, Anderson, Tattham, & Black, 1995; Tabachnick & Fidell, 2012).

Choosing a suitable factoring method for analyses is quite important for the validity of the factor structure obtained. There are various factor extraction methods like the Principal Components Analysis (PCA), Principal Axis Factoring (PAF), Maximum Likelihood Estimation (MLE), Image Factoring, Alpha Factoring, and Unweighted Least Squares and Generalized Least Squares. Among these, PCA and PAF are the most commonly used ones (Henson & Roberts, 2006; Tabachnick & Fidell, 2012). PCA is used widely than PAF is done (Thompson, 2004). The latter is usually preferred when the assumption of multivariate normality is severely violated (Costello & Osborne, 2005).

In exploratory analysis, factor rotation is applied to maximise the load of each item on a factor and to better interpret the factor structure (Rummel, 1988; Yong & Pearce, 2013). Factor rotation methods are generally divided into two. One is orthogonal, the other is oblique rotation method (Tabachnick & Fidell, 2012). Oblique rotation is used when factors are correlated; orthogonal rotation is used when factors are less correlated (Costello & Osborne, 2005; Rummel, 1988). Quartimax and Varimax are the most commonly used orthogonal rotation methods. Quartimax minimises the number of factors giving high loadings to more than one items, whereas Varimax minimises the number of items that have high loadings on more than one factor (Gorsuch, 1983; Yong & Pearce, 2013). The most widely used methods of oblique rotation include Direct Oblimin and Promax. Although both methods aim at simplifying the resultant structure, Promax is more effective in big and complicated data (Gorsuch, 1983).

Severel criteria are taken into consideration when deciding upon the number of factors (Hair et al., 1995). Some of these methods include the Kaiser criterion (eigenvalues ≥ 1), scree test, total variance explained, and parallel analysis. Factor extraction is stopped when the total variance explained reaches 95%. The precentage of total variance explained can be reduced to 50-60% in social sciences. The scree test is only reliable when you have a sample size of at least 200. A breaking point should be determined for the test (Cattell, 1978). To determine that point, researchers draw a horizontal and vertical line starting from each end of the curve (Yong & Pearce, 2013). According to Costello and Osborne (2005), a crossloading occurs when the loads of an item are at 0.32 or higher on two or more factors. If the crossloading results from the latent nature of the item, it can be retained. If it it is not so, the item may be dropped from the model or compelled to have a difference (generally 0.1) between the factor loadings (Yong & Pearce, 2013). Theoretically, there should be at least two or more items on each factor (Henson & Roberts, 2006). Factor naming is the last stage of exploratory factor analysis. Factors need to be named in consideration of what represent them best, though there is not any definite rule at this stage (Yong & Pearce, 2013).

As for *confirmatory factor analysis* (CFA), observing the adequacy of sample size, checking the assumptions, choosing the parameter estimation method, track diagram, *t* values, factor loadings, χ^2, χ^2/ degree of freedom (df), and fit indices are taken into consideration. The commony reported fit indices are as follows:

- Root Mean Square Error of Approximation (RMSEA)
- Root Mean Square Residual (RMR)
- Standardized Root Mean Square Residual (SRMR)
- Normed Fit Index (NFI)
- Non-normed Fit Index (NNFI) or Tucker Lewis Index (TLI)
- Comparative Fit Index (CFI)
- Goodness of Fit Index (GFI)
- Adjusted Goodness of Fit Index (AGFI)
- Incremental Fit Index (IFI)
- Relative Fit Index (RFI)
- Parsimonious Normed Fit Index (PNFI)
- Parsimonious Goodness of Fit Index (PGFI)

For the CFA conducting research, Kline (2016) recommends a report of RMSEA and confidence interval of 90%, χ^2, degree of freedom, as well as values of significance, CFI and SRMR. According to Cabrera-Nguyen (2010), researchers should define the model accurately, determine its constant and variable parameters, use more than one fit indices, give factor loads in tandem with significance values, explain the reasons of modifications and choose an appropriate parameter estimation method. What need to be reported in a CFA research, according to Brown (2015), include defining the model, making the data ready for analysis, model estimation, evaluation, and conclusion. Brown (2015) provides a reporting template as well.

The necessary assumptions for CFA differ from a parameter estimation method to another. Maximum Likelihood (ML) and Least Squares (LS) are the commonly preferred methods considering the data structure. Continous variables, a large sample size, and multivariate normal distribution of variables are among the assumptions of ML method (Brown & Moore, 2012). If those assumptions are violated, significance levels and goodness of model fit indices will not be estimated accurately because of the resultant model errors. LS methods are used when variables are categorical. Robust ML (Bentler, 1995) or weighted LS (Muthén, du Toit, & Spisic, 1997) methods are preferred when normality is violated with poor and moderate sample sizes. ML is widely used despite its sensitivity to sample size (Brown & Moore, 2012). It is necessary, therefore, to select a large sample as far as possible in order to

prevent the model errors. Since the adequate sample size for CFA varies in accordance with different conditions (parameter estimation method, item number, factor number), there is not an accepted number regarding the sample size (MacCallum, Widaman, Preacher, & Hong, 2001; Wolf, Harrington, Clark, & Miller, 2013), but a sample size of 300 and more is expected in general (Worthington & Whittaker, 2006). There are discussions over using the same sample with EFA and CFA. Worthington and Whittaker (2006) argue that the data structure can be set forth empirically when EFA and CFA are conducted on the same sample. However, Fabrigar, Wegener, MacCallum, and Strahan (1999) suggest a random division of the sample into two parts if it is large enough and applying EFA to one and CFA to the other. When the adequate sample size is ensured, each analysis can be conducted on different samples randomly selected from the same population. Before CFA, besides the adequate sample size, researchers should determine the imperfect and extreme values in the data, examine the uni- and multivariate normality, and check the multicollinearity (Ullman, 2012). In this way, they can both make the data ready for analysis and select a parameter estimation method suitable for the data.

It is recommended to give the factor loadings together with significance levels and to show the measurement models graphically in CFA conducting studies (Cabrera-Nguyen, 2010). Several analysis programs (MPlus, EQS, LISREL, etc.) provide these values and graphs in their outputs. χ^2 statistics is a model assessment criterion provided in outputs and frequently reported in a great deal of studies. Significance of this statistics means that the model does not fit well to the data. Although it is reported routinely in many studies, χ^2 value is usually not taken into consideration because the distribution calculated may not reflect χ^2 distribution, the value may prove to be significant in cases of large samples, and it is based on the assumption that the estimated and expected model is equal. So the model fit is decided in line with other fit indices (Brown, 2015). Due to this feature of χ^2 statistics, it can be said that χ^2/sd value is also not a good criterion for model fit. Even though some researchers (Anderson & Gerbing, 1984) consider χ^2/sd as an acceptable criterion for model fit if it is smaller than 5, Kline (2016) points out that using this value as a criterion for model fit does not have logical and statistical grounds. In the same vein, Wheaton (1987) suggested not to use this value in model fit assessments.

There are a number of indices used to assess model fit and various opinions on which indices need to be reported. Kline (2016) suggests researchers to report RMSEA and confidence interval of 90%, χ^2, CFI and SRMR values at least in CFA studies, while Cabrera-Nguyen (2010) proposes the use of more than one fit indices. According to Brown (2015), fit indices are divided into three groups, videlicet the absolute fit indices (χ^2, SRMR ve RMR), parsimonious indices (RMSEA), and comparative indices (CFI-IFI, TLI-NNFI); at least one index from each group should be used in reporting. Hu and Bentler (1999) propose using the couples of NNFI (min. 0.96) and SRMR (max. 0.09), or RMSEA (min. 0.06) and SRMR (max. 0.09), or CFI min. 0.96) and SRMR (max. 0.09) in line with the two-index strategy. Although GFI and AGFI are frequenlty used, they are not recommended due to their poor performance in simulation studies (Hu & Bentler, 1998; Marsh, Balla, & McDonald, 1988). Crowley and Fan (1997) propose reporting as many indices as possible, because each index gives information about a different aspect of the model fit. It is not appropriate to line up definite criteria for fit indices as they are affected by a range of factors such as sample size, model complication level, estimation method (e.g. ML and LS), data type (continuous and categorical), normality of data, misdefinition of model, etc. (Brown, 2015). Which indices should be used under what conditions when assessing the model fit can be found in the relevant literature (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999; Kline, 2016; MacCallum, Browne, & Sugawara, 1996; Marsh & Hau, 1996; Steiger, 2007; Tabachnick & Fidell, 2012; Wheaton, 1987). The acceptable values for fit indices compiled from the related literature are given in Table 1.

Table 1. Fit Indices and Acceptable Cut-off Points

Statistics	Interpretation
χ^2/sd	Should be above 5 (Anderson & Gerbing, 1984); Not recommended (Kline, 2016; Wheaton,1987)
RMSEA	Should be close to or below 0.06 (Hu & Bentler, 1999); Below 0.08 is acceptable, below 0.05 shows a good fit, model rejected if above 0.1 (Browne & Cudeck, 1993); Between 0.08 to 0.10 is mediocre fit (MacCallum et al., 1996); Should be close to or below 0.07 (Steiger, 2007)
RMR	Should be as low as possible (Tabachnick & Fidell, 2012); Not preferred due to its sensitivity to scale level (Hooper et al., 2008)
SRMR	Should be close to or below 0.08 (Hu & Bentler, 1999); Below 0.05 means a good fit (Byrne, 1998)
NFI	Above 0.90 means a good fit (Bentler & Bonnet, 1980); Should be above 0.95 (Hu & Bentler, 1999); Up to 0.80 is acceptable (Hooper et al., 2008)
NNFI-TLI	Should be close to or above 0.95, between 0.90 and 0.95 is acceptable, model error if below 0.90 (Hu & Bentler, 1998, 1999)
CFI	Close to or above 0.95 shows a good fit, 0.90-0.95 is acceptable, model error if below 0.90 (Hu and Bentler, 1998, 1999)
IFI	0.90 and above shows a good fit (Marsh & Hau, 1996)
RFI	0.90 and above shows a good fit (Marsh & Hau, 1996)
GFI	Should be above 0.95 (Miles & Shevlin, 1998); 0.90 and above shows a good fit (Hooper et al., 2008)
AGFI	0.90 and above shows a good fit (Hooper et al., 2008)
PNFI	There is not a certain limit but can be as low as 0.50 (Hooper et al., 2008; Mulaik et al., 1989); Should be above 0.80 (Marsh & Hau, 1996)
PGFI	There is not a certain limit but can be as low as 0.50 (Hooper et al., 2008; Mulaik et al., 1989); Should be above 0.80 (Marsh & Hau, 1996)

Since the fit index values given in Table 1 are examined, it is recommended that the RMSEA value should be less than 0.10 and SRMR should not be less than 0.80. Considering the goodness of fit indexes, it is recommended that PNFI and PGFI values should be greater than 0.50, NFI should be greater than 0.80 and other index values should be greater than 0.90.

Hitherto we dealt with the basic procedure of output interpretation and necessary conditions of factor analyses that provide significant evidence for the construct validity in scale development research. This paper aims to examine some research in the related literature in terms of factor analysis processes and, in this way, to achieve considerable results on how EFA and CFA, which are frequenlty referred in theory, are applied in practice. Within the scope of this general aim, we noticed that there is a limited number of studies on scale development and adaptation (Acar-Güvendir & Özer-Özkan, 2015; Bayık & Gürbüz, 2016; Çüm & Koç, 2013; Delice & Ergene, 2015; Erkuş, 2016; Gül & Sözbilir, 2015; Şahin & Boztunç Öztürk, 2018; Öztürk, Eroğlu, & Kelecioğlu, 2015) and factor analysis processes

(Thompson & Daniel, 1996; Worthington & Whittaker, 2006). Those studies deal with papers in Social Science Citation Index (SSCI), a research sample, or how they handle analysis processes. The present paper is important, therefore, in terms of examining the factor analysis processes used in all of the scale development studies published between 2006 and 2016 in TUBİTAK ULAKBİM National Data Base. Reviewing all of the scale development papers published in the field of social sciences in the last decade in TUBİTAK ULAKBİM that is the most inclusive research database in Turkey is thought to light the way for further research on the use of EFA and CFA. The research questions are as follows: In the scale development studies published between 2006 and 2016 in Turkey;

1. What sort of results are reached regarding the use of exploratory factor analysis in the studies examined?
2. What sort of results are reached regarding the use of exploratory factor analysis in the studies examined?

Method

The present paper is a descriptive research due to the fact that it reports the existing properties of the scale development studies in reference to exploratory and confirmatory factor analyses. Descriptive studies aim at passing on the attributes of an analysis unit as they are, rather than scrutinising their reasons (Fraenkel & Wallen, 2011). Document analysis, one of the qualitative research methods, was used in the study. Document analysis is a systematic procedure to evaluate and review both printed and electronic materials. It can be used as a supplemental research method or as a separate technique (Bowen, 2009).

Study Group

This study examined totally 121 scale development studies published in the fields of education and social sciences in the journals indexed in TUBİTAK ULAKBİM Dergi Park data base between 2006 and 2016. The distribution of those studies according to years are given in Figure 1.

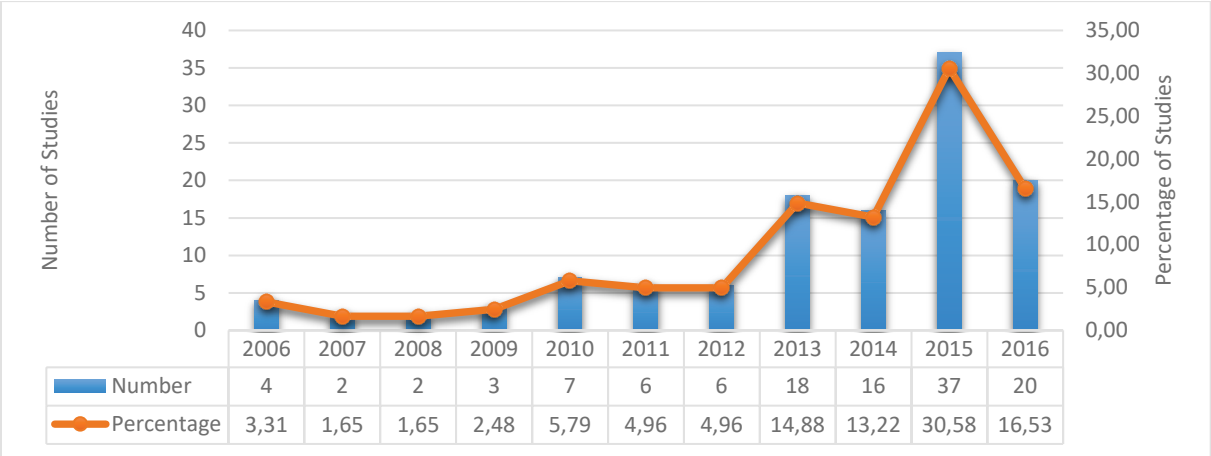


Figure 1. Distribution of Studies to Years

As seen in Figure 1, the number of scales developed considerably increased in progress of time. More than 70% of the scale development studies examined in this paper were carried out after 2013. The maximum number of studies was published in 2015, and the minimum number of studies was published in 2008 (Figure 1). EFA was used in 97% (N_{EFA}=117) of those studies and CFA was used in 51% (N_{CFA}=62). Both analyses were employed in 48% (N_{BOTH}=58) of the studies. EFA and CFA were conducted on the same sample in 62% (N_{SAME}=36) and on different samples in 38% ($N_{DIFFERENT}$=36) of 58 studies that used both EFA and CFA.

Procedure

Within the scope of the research, we excogitated firstly on which data base or bases to take into consideration. Instead of examining the studies only in a certain index, we decided to examine the publications in the field of Social Sciences in TÜBİTAK Ulakbim National Data Base which is the most inclusive one in Turkey. The keywords searched on the data base were "psychometric", "scale", "factor analysis", "validity", and "development". Following the search, 10 out of 131 studies published between 2006 and 2016 were excluded from the research since they did not involve EFA and/or CFA in their analyses (Appendix 1). The search and selection of papers were followed by a detailed examination of the primary and secondary sources in the literature on what is required to conduct EFA and CFA and what are the possible outputs needing to be reported. What were taken into consideration for EFA include the sample size, factor extraction method, Kaiser-Meyer-Olkin (KMO) and Bartlett tests, whether scree plot and factor loadings were examined or not, factor rotation method, factor naming, and whether the percentage of total variance explained was reported. What we examined for CFA include the sample size, whether the assumptions were checked, what parameter estimation method was used, whether track diagram, t values, factor loads and χ^2 significance values were reported, χ^2/sd, as well as RMSEA, RMR, SRMR, NFI, NNFI, CFI, GFI, AGFI, IFI, TLI, RFI, and PGFI. As for the studies involving both analyses, it was seen important to examine whether confirmatory factor analysis was conducted on the same sample or not. The data were input via Microsoft Office Excel 2016 after determining what criteria to take into consideration prior to examining the studies.

Data Analysis

Document analysis technique which is in the scope of the present research, has advantages such as easy access to many documents, no need of interaction with individuals or individuals, wide time period for the research, easy access to large sample size, individuality and authenticity, low cost and access to qualified resources. Some biases of findings, access to only very well-known or recognized works, obtaining indirect information because of lack of real information in some cases, the difficulty of access to resources, the adequate sample and coding difficulties can be considered as the main deficiencies of document analysis technique (Bailey, 1982). Document review consists of reaching the documents, examining and understanding the originality of documents, analyzing and reporting the data steps. In this study, after reaching the articles in the education and social sciences fields published in the TR Index, the data in the publications were examined and reported with appropriate statistical and graphical methods.

Following the entry to Microsoft Office Excel 2016 of the data reported for EFA and CFA in the scale development studies, we provided line charts with clustered columns and markers related to the distribution of studies to years, the number of studies using different types of analysis, as well as the number and percentage of studies conducting the two analyses on the same sample. Line charts with clustered columns and markers, frequencies and percentages of factor rotation methods, total variance explained, factor numbers, factor extraction methods, and average sample size for the studies with EFA are presented. In addition, frequencies and percentages on whether Kaiser-Meyer-Olkin (KMO) and Bartlett tests, scree plot and factor loads were examined and whether factors were named or not are presented in tables. As for the CFA conducting studies, line charts with clustered columns and markers, frequencies and percentages about the average sample size, parameter estimation method used, χ^2/sd and the average fit index are presented. The frequencies and percentages on whether the assumptions were checked, whether track diagram, t values, factor loads and χ^2 significance values were reported are presented in tables as well. The offset and threshold values in Table 1 were used when interpreting the fit index values.

In the coding phase of the document analysis, 10 articles were selected from 121 articles by two researchers in order to ensure reliability. In order to ensure intercoder reliability, the reliability formula proposed by Miles and Huberman (1994) (Number of agreed views / (Number of agreed views + Number of disagreed views)) was used and the agreement coefficient was calculated as 0.92. In terms of consistency between the data in the process of reviewing the articles, the researchers exchanged ideas during the process where necessary.

Results

The results of the research need to be explained in the relevant tables, figures, charts, and images in a way to support the aim and problem of the research as well as to ensure the integrity thereof.

The Results on the Use of Exploratory Factor Analysis

The distribution of the EFA using studies according to their sample sizes are presented in Figure 2.

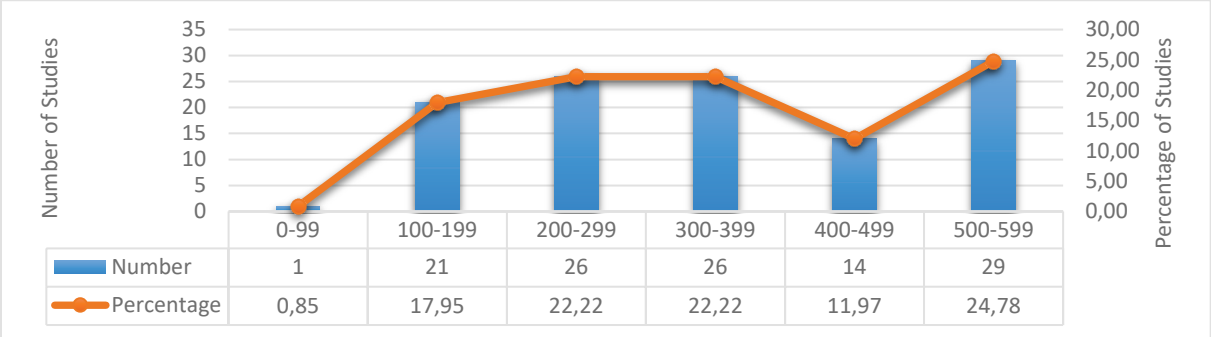


Figure 2. Distribution of EFA Using Studies According to Sample Sizes

As seen in Figure 2, the most of the studies were carried out with a sample size over 500 (n=29). The average sample is aproximately 395 for all of the studies. The number of studies with an average sample size of 300 and above is more than half of the total number of studies (n=69). It can be said accordingly that the majority of the studies were conducted with samples greater than 300.

The distribution of EFA using studies according to the factor extraction method they used is given in Figure 3.

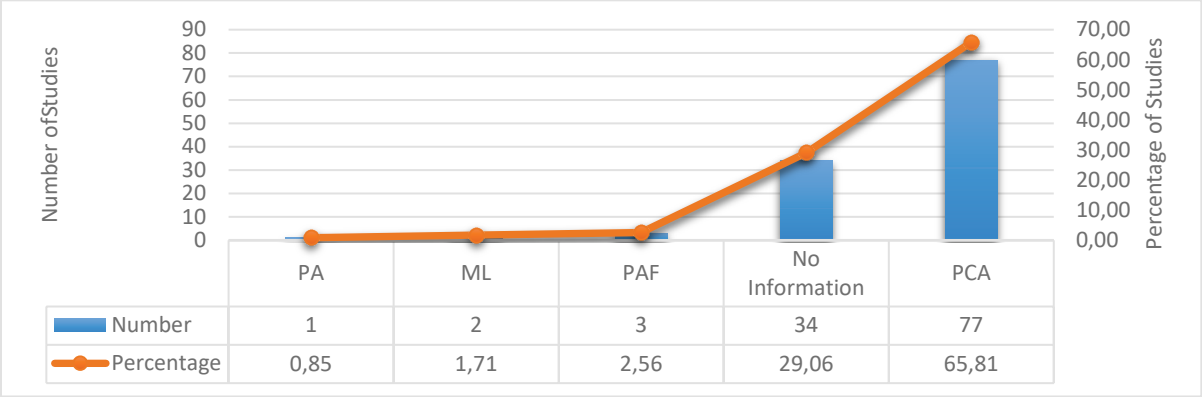
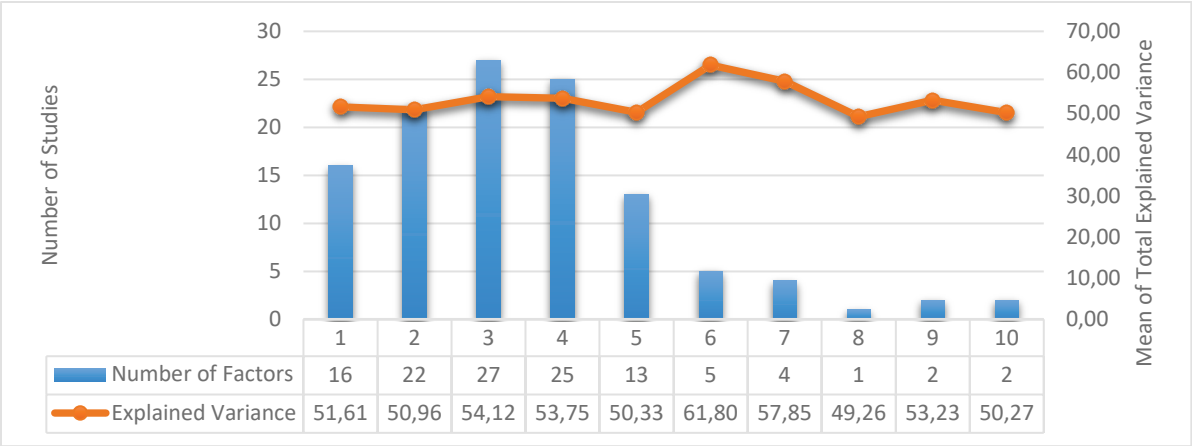


Figure 3. Distribution of Studies According to Factor Extraction Methods They Used

As seen in Figure 3, a clear majority (%65.81) of the studies used the PCA factor extraction method. Another remarkable finding is that more than a quarter (%29.06) of the studies did not give any information on factor extraction method. 3 out of the studies we examined used PAF and 2 used ML method. 1 study used Parallel Analysis (PA) as factor determination method.

Figure 4 shows the distribution of EFA using studies according to the number of factors obtained and the percentage of total variance explained.

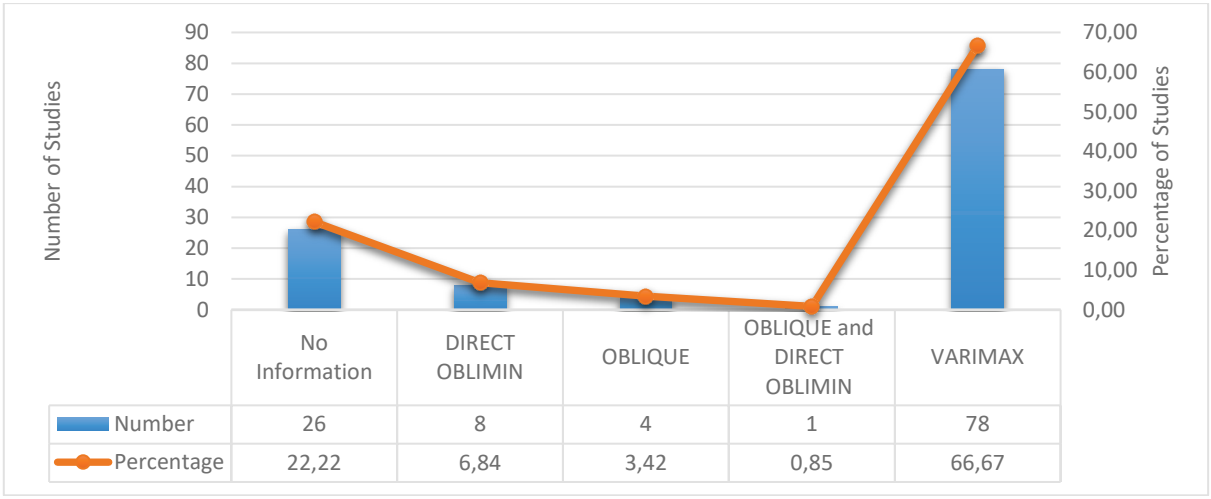


	1	2	3	4	5	6	7	8	9	10
Number of Factors	16	22	27	25	13	5	4	1	2	2
Explained Variance	51,61	50,96	54,12	53,75	50,33	61,80	57,85	49,26	53,23	50,27

Figure 4. Distribution of Studies According to the Number of Factors Obtained and the Percentage of Total Variance Explained

As seen in Figure 4, 13.68% of the studies were with one dimension. A large majority of studies (*n*=103) ranges from 1 to 5 factors, while the number of studies (n=27) with three dimensions is the most. The percentage of average total variance explained for the one-factor structures is 51.61, whereas it varies from 49% to 62% for the structures with two or more factors. The percentage of average total variance explained did not increase considerably in tandem with the increase in number of factors.

Figure 5 shows the distribution of EFA using studies according to the factor rotation methods they used.



	No Information	DIRECT OBLIMIN	OBLIQUE	OBLIQUE and DIRECT OBLIMIN	VARIMAX
Number	26	8	4	1	78
Percentage	22,22	6,84	3,42	0,85	66,67

Figure 5. Number of Studies Distributed According to Factor Rotation Methods

As Figure 5 demonstrates, 66.67% of the studies used VARIMAX rotation method. 22.22% of the studies did not give any information on their factor rotation method. OBLIQUE and DIRECT OBLIMIN were reported in 4 and 8 papers, respectively. One paper states that it used both of the oblique rotation methods.

Table 2 shows the data on whether KMO and Bartlettt values, scree plot, factor naming and factor loadings were reported or not in EFA using papers examined in this research.

Table 2. Reporting the Values of KMO and Bartlett, Scree Plot, Factor Naming and Factor Loadings

	KMO Value		Bartlett Test		Scree Plot		Factor Naming		Factor Loading	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Yes	110	94.02	109	93.16	58	49.60	93	91.18	112	95.73
No	7	5.98	8	6.84	59	50.40	9	8.82	5	4.27
Total	117	100	117	100	117	100.0	102	100.0	117	100.0

As seen in Table 2, KMO and Bartlett statistics were reported respectively in 94.02% and 93.16% of 117 EFA using studies. While scree plot was reported in 49.60% of the studies, factor loadings were given in 95.73% of the papers. Factors were named in 91.18% of the papers, whereas the remaining 8.82% did not fulfill factor naming.

Findings on the Use of Confirmatory Factor Analysis

Table 3 shows whether 62 CFA using studies checked their assumptions, whether they demonstrated their measurement model with figures, and whether they reported *t* values, factor loadings, and χ^2 significance values.

Table 3. Checking the Assumptions, Figural Demonstration of the Measurement Model, Reporting t Values, Factor Loadings, and χ2 Significance Values

	Checking Assumptions		χ^2 Significance Value		Figural Demonstration of Measurement Model		*t* values		Factor Loadings	
	n	%	n	%	n	%	n	%	n	%
Yes	8	12.90	46	74.19	51	82.26	16	25.81	52	83.87
No	54	87.10	16	25.81	11	17.74	46	74.19	10	16.13
Total	62	100.0	62	100.0	62	100.0	62	100.0	62	100.0

As seen in Table 3, 87.10% of 62 CFA using papers did not check the assumptions required for the analysis. Even if the assumptions were checked, they were not reported. 82.26% of the studies demonstrated their measurement models through figures. Factor loadings were reported in 3 out of 11 studies that did not figuralised their measurement models. The remaining 8 studies did not report the factor loadings. Factor loadings were given in 83.87% of the studies. χ^2 significance values were given in 74.19% of the studies, while *t* values were reported in 25.81%.

The distribution of the CFA using studies according to their sample sizes is given in Figure 6.

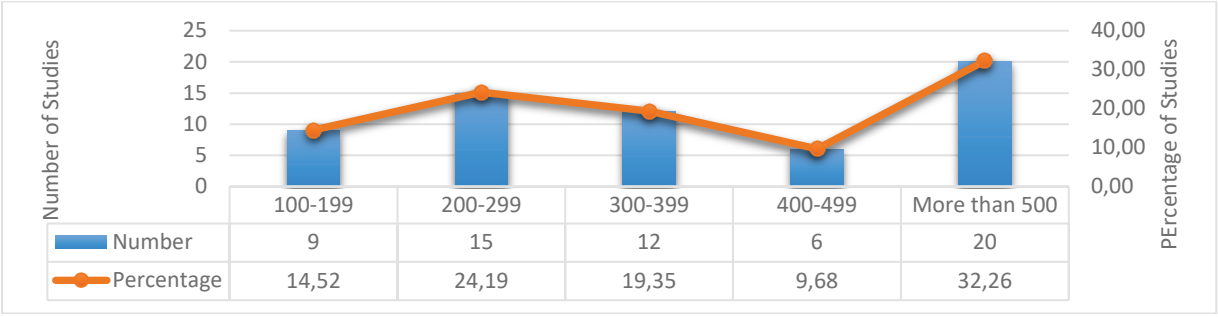


Figure 6. Distribution of CFA Using Studies According to Their Sample Size Ranges

When Figure 7 is examined, it is seen that 38.71% of the CFA using studies were conducted with a sample size below 300. When the CFA using studies are viewed with respect to their sample sizes, it is seen that the majority of the studies were carried out with 415 participants on an average. While 61.29% of the studies had a sample size of 300 and above, the maximum number of research was conducted with a sample size of 500 and above. Hence, it can be said that the studies carried out with exiguous samples are relatively few in number.

Figure 7 shows the distribution of the CFA using studies according to the parameter estimation method they used.

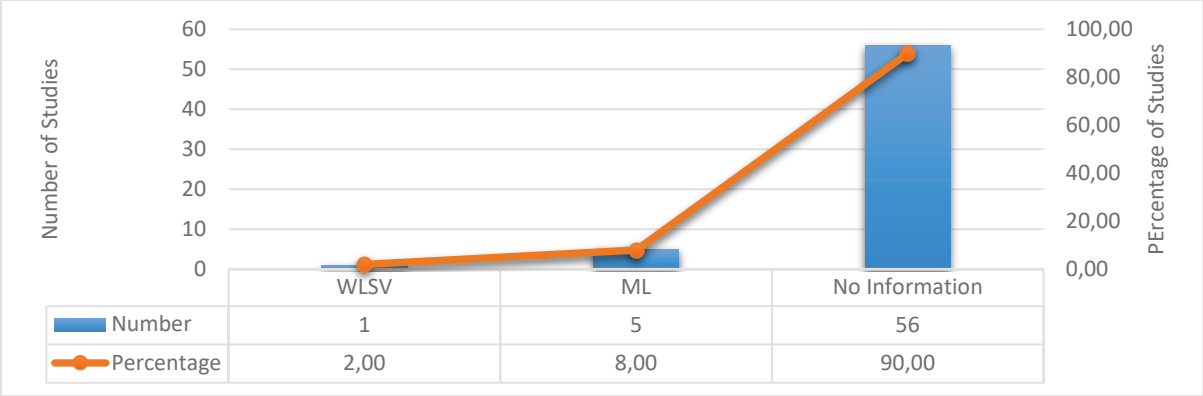


Figure 7. Distribution of the Number of Studies According to Their Parameter Estimation Method

As seen in Figure 7, 90% of the studies did not give any information about their parameter estimation method. ML and WLSV were used in respectively 8% and 1% of the studies. Apart from the mentiones ones, any other parameter estimation method was not reported.

All comments regarding the adequacy of the fit index values used in the studies that conducted CFA were made according to the criteria given in Table 1.

Figure 8 shows the distribution of CFA using studies according to χ^2/sd ratio as a fit statistic.

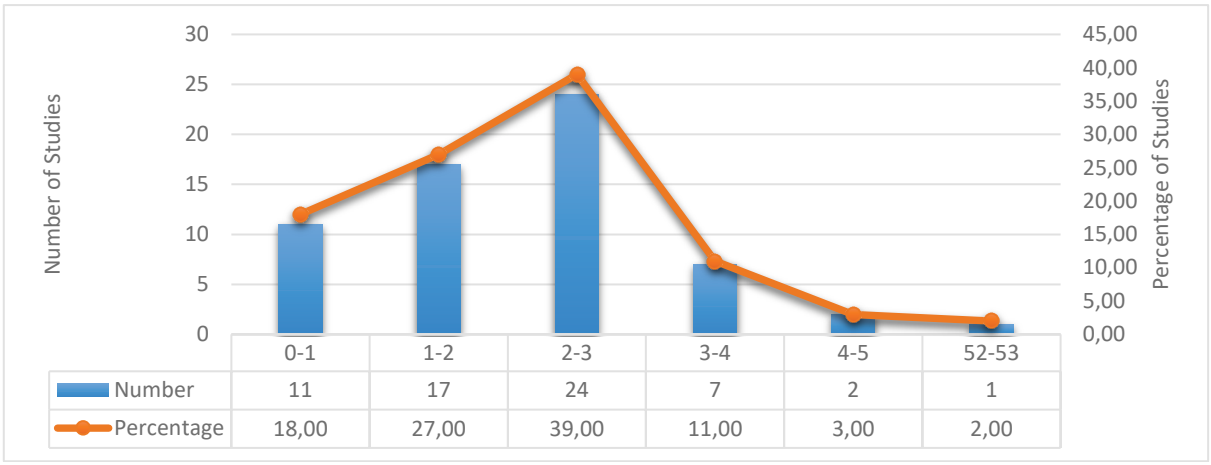


Figure 8. Distribution of Studies According to χ^2/sd Ratio Ranges

As indicated in Figure 8, χ^2/sd ratio was reported in all of the CFA using studies. Except for one CFA study, χ^2/sd ratio was at an acceptable range in all of them. The ratio was between 2 and 3 in most of the studies, but it was below 3 in general when all of the studies are considered. As the use of χ^2/sd ratio as a criterion of model fitting is arguable, per study interpretation of the other fit indices is reported below.

Figure 9 presents the distribution of how many fit indices were used by each study examined in this research.

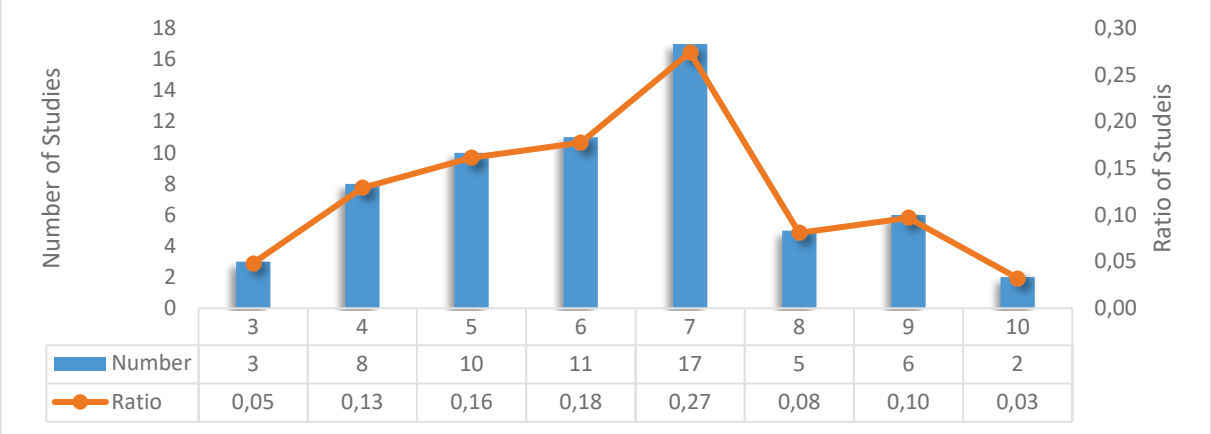


Figure 9. Number of Fit Indices Used Per Study

As seen in Figure 9, the studies conducted with CFA used 3 to 10 indices to assess the model data fitting and each study used at least 3 fit indices. Model fitting was determined via 6 or more indices in 66.13% of the studies.

The distribution of the CFA using studies according to the RMSEA value ranges is given in Figure 10.

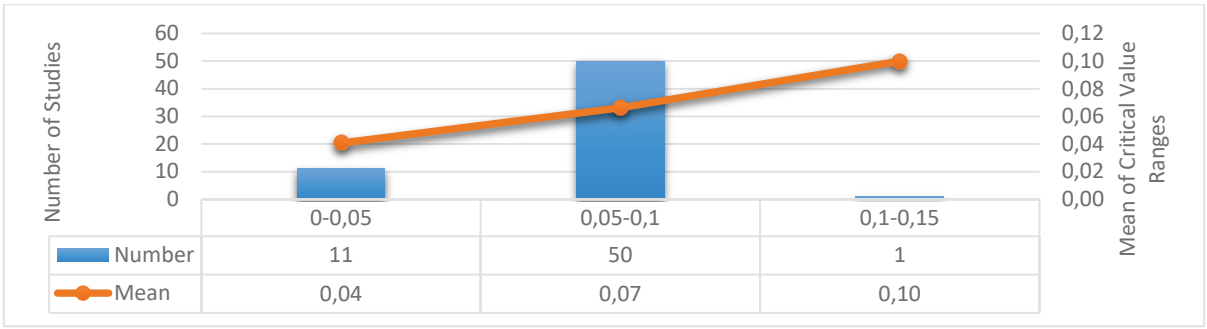
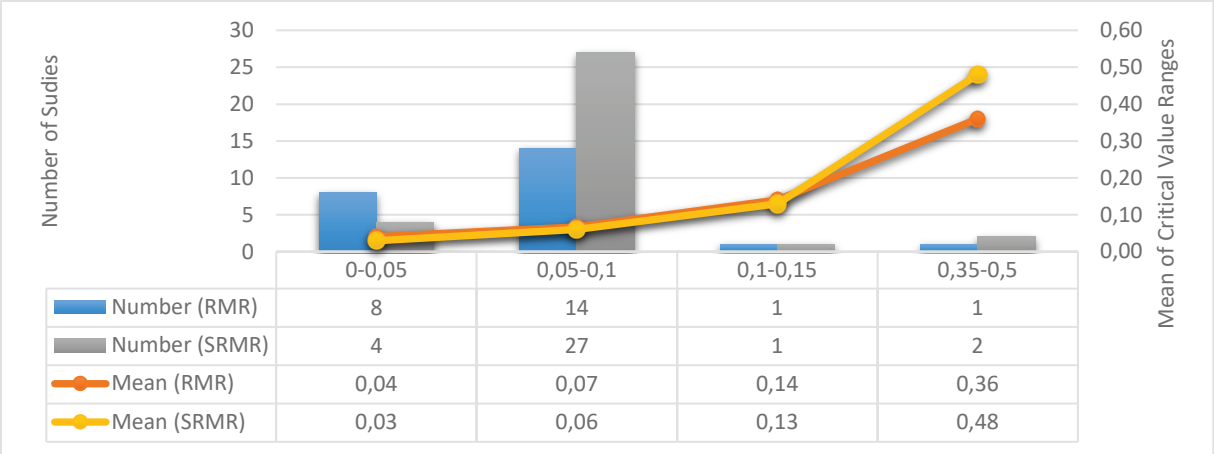


Figure 10. Number and Average of Studies for Critical RMSEA Value Ranges

When Figure 10 is examined, it is seen that the RMSEA fit index was reported in every CFA using research. While 17.74% of the papers showed an excellent fit range, 80.65% had the RMSEA values with an acceptable range. A small percentage of the papers showed a RMSEA value out of the fit ranges. It can be concluded that the RMSEA values supported the model data fitting in 98.39% of the papers examined in this research.

The distribution of the CFA using studies according to the RMR and SRMR value ranges is shown in Figure 11.

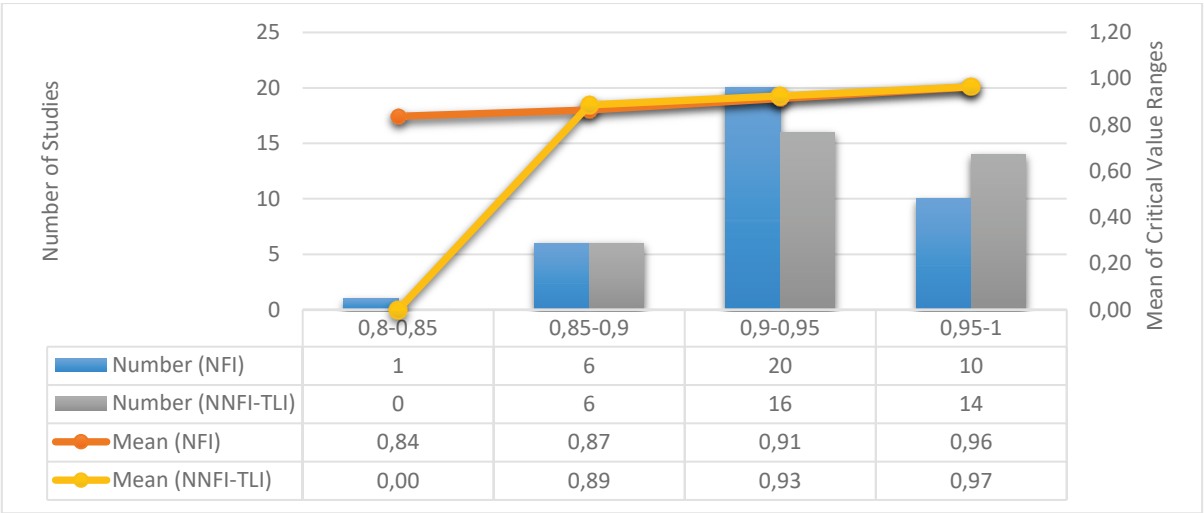


	0-0,05	0,05-0,1	0,1-0,15	0,35-0,5
Number (RMR)	8	14	1	1
Number (SRMR)	4	27	1	2
Mean (RMR)	0,04	0,07	0,14	0,36
Mean (SRMR)	0,03	0,06	0,13	0,48

Figure 11. Number and Average of Studies for Critical RMR and SRMR Value Ranges

In line with Figure 11, it is seen that the RMR fit statistics was reported in 24 (38.71%) out of 62 CFA using studies. While 33.33% of the studies reporting the RMR statistics showed a perfect fit range, 58.33% had RMR values within an acceptable range. This value was out of the fit values in 8.33% of the papers in which RMR statistics was reported. It shows that they did not provide the model data fitting. Hence, it can be said that RMR values in 91.67% of the papers supported the model data fitting. The SRMR fit statistics was reported in 34 (54.84%) out of 62 CFA using studies. Whereas 11.77% of the papers reporting the SRMR statistics showed a prefect fit range, 79.41% had SRMR values within an acceptable range. This value was out of the fit values in 8.82% of the papers reporting SRMR statistics, which shows that they could not attain the model data fitting. It can be interpreted accordingly that SRMR values in 91.18% of the papers supported the model data fit.

Figure 12 shows the distribution of the CFA using studies according to NFI and NNFI-TLI value ranges.



	0,8-0,85	0,85-0,9	0,9-0,95	0,95-1
Number (NFI)	1	6	20	10
Number (NNFI-TLI)	0	6	16	14
Mean (NFI)	0,84	0,87	0,91	0,96
Mean (NNFI-TLI)	0,00	0,89	0,93	0,97

Figure 12. Number and Average of Studies for Critical NFI and NNFI (TLI) Value Ranges

As seen in Figure 12, NFI fit statistics was reported in 37 (59.68%) out of 62 CFA using papers. 27.03% of the papers in which NFI statistics was reported showed a perfect fit range, while NFI values in 54.05% of the papers were within an acceptable fit range. This value was out of the fit values in 18.92% of the papers reporting NFI statistics, indicating that model data fitting was not ensured. It can be interpreted that NFI values in 81.08 of the papers supported the model data fitting. NNFI fit statistics was reported in 36 (58.07% out of) 62 papers in which CFA was used. 38.89% of the papers with a report of NNFI statistics had a perfect fit range, whereas NNFI values were within an acceptable range in 44.44% of the papers. This value was out of the fit values in 16.67% of the papers reporting NNFI

statistics, which shows that model data fitting was not attained. It can be said, therefore, that NNFI values in 83.33% of the papers supported the model data fitting.

Figure 13 shows the distribution of CFA using studies according to their ranges of CFI value.

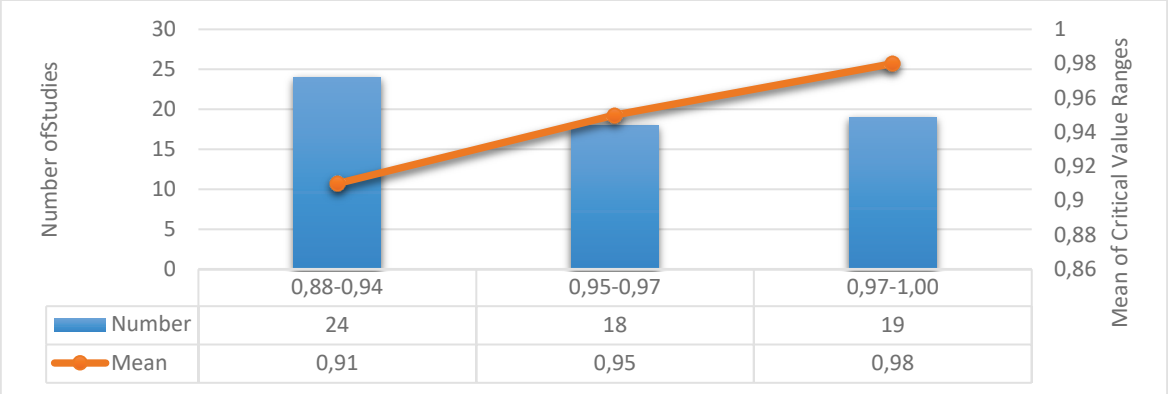


Figure 13. Number and Average of Studies for Critical CFI Value Ranges

As seen in Figure 13, CFI fit statistics was reported in 61 (98.39%) out of 62 CFA using papers. 31.15% of the papers reporting CFI statistics had a perfect fit range, whereas CFI values were within an acceptable range in 29.51% of the papers. This value was out of the fit values in 39.34% of the papers reporting NNFI statistics, which shows that model data fitting was not attained. It can be concluded that CFI values in 60.66% of the papers supported the model data fitting.

The distribution of CFA using studies according to their ranges of IFI, GFI and RFI values is presented in Figure 14.

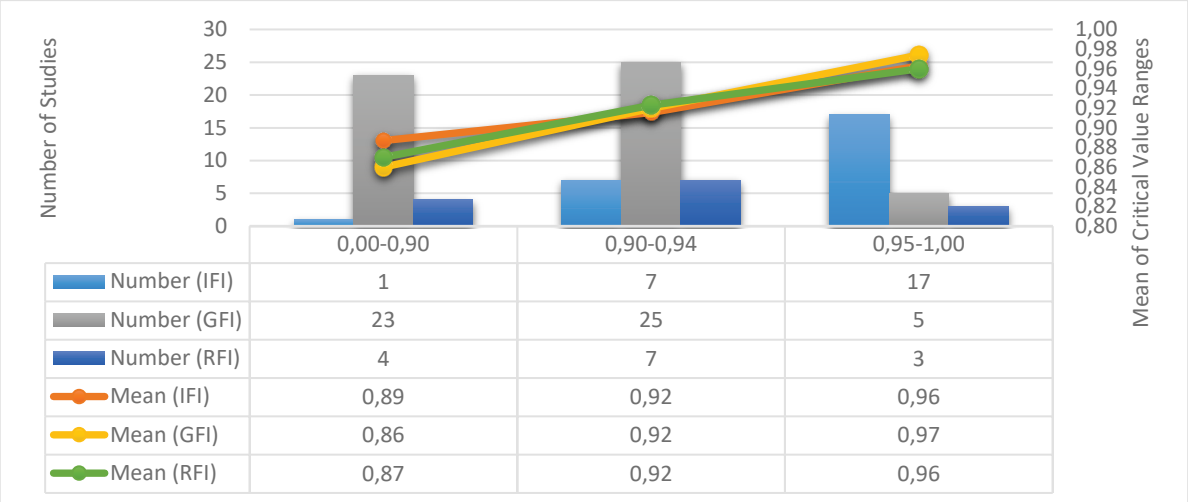


Figure 14. Number and Average of Studies for Critical IFI, GFI and RFI Value Ranges

Figure 14 shows that IFI fit statistics was reported in 25 (40.32%) out of 62 CFA using papers. 68% of the papers reporting IFI statistics showed a perfect fit range, whereas IFI values were within an acceptable range in 28% of the papers. This value was out of the fit values in 4% of the papers reporting IFI statistics, which shows that model data fitting was not attained. Therefore, it can be concluded that IFI values in 96% of the papers supported the model data fitting. GFI fit statistics was reported in 53 (85.48%) out of 62 CFA using papers. 9.43% of the papers reporting IFI statistics showed a perfect fit range, whereas GFI values were within an acceptable range in 47.17% of the papers. This value was out of the fit values in 43.40% of the papers reporting GFI statistics, which shows that model data fitting was not attained. It can be concluded that GFI values in 56.60% of the papers supported the model data fit. RFI fit statistics was reported in 14 (22.58%) out of 62 CFA using papers. 21.43% of the papers

reporting RFI statistics had a perfect fit range, while RFI values were within an acceptable range in 50% of the papers. This value was out of the fit values in 28.57% of the papers reporting RFI statistics, which shows that model data fitting was not attained. It can be said that RFI values in 71.43% of the papers supported the model data fit.

The distribution of CFA using studies according to their ranges of AGFI value is given in Figure 15.

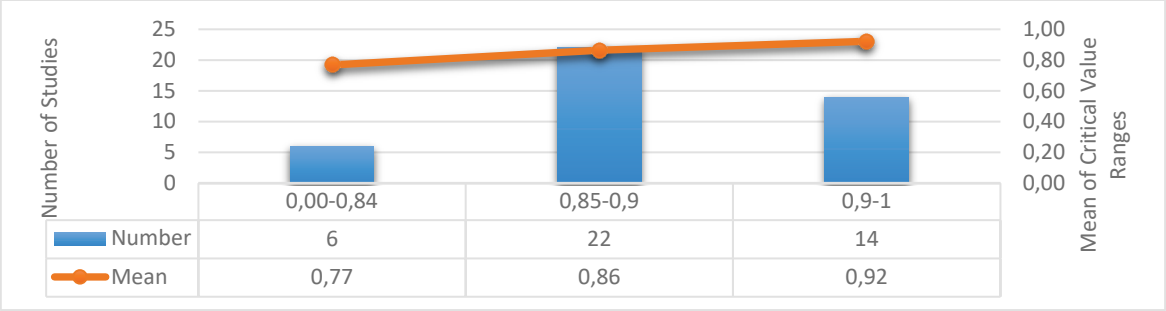


Figure 15. Number and Average of Studies for Critical AGFI Value Ranges

As seen in Figure 15, AGFI fit statistics was reported in 42 (67.74%) out of 62 CFA using papers. 33.33% of the papers reporting AGFI statistics had a perfect fit range, while AGFI values were within an acceptable range in 52.38% of the papers. This value was out of the fit values in 14.29% of the papers reporting AGFI statistics, which shows that model data fitting was not attained. It can be said that AGFI values in 61.91% of the papers supported the model data fit.

PNFI fit statistics was not applied in any papers examined in this study, whereas PGFI was reported in 4 (6.45%) papers. Reported values of PGFI ranges from 0.65 to 0.80.

Discussion, Conclusion and Suggestions

This study investigated the use of exploratory and confirmatory factor analyses in the scale development studies published in the fields of education and social sciences in Turkey between 2006 and 2016. The majority of the studies examined were carried out after 2013. The fact that a large majority of the studies used EFA and nearly half of the studies used CFA indicates that some studies did not need to hypothetically verify the scale factor structures they reached. Factor analysis provides evidence for the construct validity of a scale (Cronbach & Meehl, 1955). Testing the factor structures with CFA provides stronger evidence for construct validity. More than half of the studies that used both analyses grounded their analyses on the same sample. According to Worthington and Whittaker (2006), carrying out analyses over the same sample is not problematic, but Fabrigar et al. (1999) suggests conducting analyses on two randomly divided samples if the sample is large enough. However, the real factor structure may not be put forth in CFA, because item sequence is likely to change when too many items are excluded from the draft scale following the exploratory analysis, as well as some problems (weariness, confusion, interaction, etc.) may affect the way participants respond to questions. Therefore, obtaining a factor structure through an exploratory analysis and testing that structure hypothetically over a sample (randomly selected from the target population) with similar characteristics will provide more valid and reliable a priori evidence for scale development.

The studies conducted with EFA had an average sample size above 300, and 60% of the studies were conducted with more than 300 participants. Sample size may change according to item numbers and factor loading magnitudes (Guadagnoli & Velicer, 1988; Streiner, 1994; Yong & Pearce, 2013). At least a sample size of 300 is recommended for EFA (Comrey & Lee, 1992; Tabachnick & Fidell, 2012). Sample sizes below 300 are considered moderate or inadequate (Comrey & Lee, 1992). It can be said that adequate sample sizes were provided in most of the studies examined in this research. However, moderate or undersized samples were also employed in a considerable number of studies. In the same way, the CFA using studies had an average sample size above 300, and nearly 60% of the studies were

conducted with more than 300 participants. A sample size of 300 and above is expected for CFA in general (Worthington & Whittaker, 2006), though the adequate size may vary according to different conditions (MacCallum et al., 2001; Wolf et al., 2013). Adequate sample sizes are reached in most of the studies conducted with CFA, but a considerable amount of studies had a small sample size. When scale development studies are not carried out with a sufficient number of participants, factor structures obtained and results inferred from scores of that scale may be erroneous. It is recommended, therefore, to consider at least two of the criteria in the literature when determining the adequate sample size prior to carrying out a study (Çokluk, Şekercioğlu, & Büyüköztürk, 2012).

PCA factor extraction method was used in a large majority of the studies, but they did not give any explanation about why they preferred that method. This makes us think that PAF, MLE, and PA could be used as well in accordance with the data structure. What is assumed methodically is a determinant factor for the method to be used in factor analysis. For example, PAF is used when multivariate normality is violated (Costello & Osborne, 2005). In a similar way, VARIMAX rotation method was used in a large majority of the studies without any explanation about why that method was preferred. This makes us think that the other orthogonal and oblique rotation methods could be used as well in accordance with the data structure. Oblique rotation is used when factors are correlated, while orthogonal rotation is used when factors are less correlated (Costello & Osborne, 2005; Rummel, 1988). Orthogonal and oblique rotation methods are preferred in line with different conditions (Gorsuch, 1983; Yong & Pearce, 2013). More than a quarter of the studies examined in this research did not mention which factor extraction and rotation methods they used. This raises a question mark about how they obtained their scale structures. Likewise, analysis methods and assumptions related to those methods were not examined in a large majority of the studies conducted with confirmatory analysis. Choosing an appropriate analysis method is among the crucial steps of confirmatory factor analysis (Brown, 2015; Cabrera & Nguyen, 2010). Additionally, prior to conducting CFA, it is recommended to examine multivariate normality and multicollinearity, as well as missing and extreme values (Ullman, 2012). In conclusion, clearly stating the reasons for opting a method of factor extraction, factor rotation, and parameter estimation to be used in exploratory and confirmatory factor analysis will possibly remove the question marks about how the factor structures were obtained and how they were tested.

Item factor loadings were reported in a vast majority of the studies in both analysis methods. In exploratory analysis, whether the data used fit to the analysis (KMO and Bartlet tests) was examined and the factors were named. However, scree plots were given only in half of the studies. Scree plot is one of the criteria taken into consideration when deciding the factor number (Hair et al., 1995; Çokluk et al., 2012; Yong & Pearce, 2013). Giving place to scree plot that is frequently included in outputs of exploratory analysis when determining the factor structure will provide evidence for the factor structure. While the studies examined were mostly made up of more than one factor, the total variance explained was between 50 and 60%. Total variance explained usually remains at 50-60% in the fields of education and social sciences (Yong & Pearce, 2013). Total variance explained is recommended to be over 30% at least for single factoral structures (Büyüköztürk, 2013). It can be said that the relevant percentages were accordingly sufficient in the papers we examined. However, if the percentage of total variance explained does not increase in tandem with the number of factors, it means that the percentage of total variance explained per factor decreases and throws suspicion on the extent to which factors represent the psychological structure. In such cases, it is useful to evaluate the factor structures together with the measured structure, rather than considering them separately.

Giving χ^2 significance values of the measurement model and the figural depiction thereof makes it easier to examine and comprehend the models. Nevertheless, *t* significance values of the factor loadings were not given in most of the studies. In CFA studies, model estimations including the significance values of factor loadings are recommended (Brown, 2015; Cabrera & Nguyen, 2010). Accordingly giving the item significance values on a figural demonstration of the measurement model is seen important in terms of assessing the factor loadings in confirmatory analyses.

χ^2/sd values were reported in all of the CFA using studies examined and the reported values were below 5 in almost all of them. Whereas some researchers (Anderson & Gerbing, 1984) consider χ^2/sd below 5 as an acceptable criterion of model fit, others (Kline, 2016; Wheaton, 1987) points out that using this value as a criterion of model fit does not have any logical and statistical grounds. χ^2/sd values were at an acceptable level in almost all of the studies we examined, but the fact that a disputable index was reported in all of the studies is a drawback needing to be addressed. Although there is a great deal of indices to be used to test the model fit, regarding a value without a sufficient logical and statistical basis as a criterion of model fit is an undesirable situation.

The values of RMSEA and CFI fit indices were reported in almost all of the papers examined. At least three fit indices were used in all of the studies and more than five indices were used in a vast majority of the studies. The other most used indices are GFI, AGFI, NFI, NNFI-TLI, SRMR, IFI, RMR, RFI, and PGFI, respectively. In conducting CFA, some researchers (Brown, 2015; Hu & Bentler, 1999; Kline, 2016) suggest certain fit indices (χ^2, CFI-IFI, TLI-NNFI, SRMR-RMR, RMSEA), while others suggest using a good number of fit indices as far as possible (Cabrera-Nguyen, 2010; Crowley & Fan, 1997). It can be said accordingly that an adequate number of model fit indices were used in almost all of the studies conducted with CFA. Although the GFI and AGFI fit indices were used more frequently than the other indices, they are not recommended due to their poor performance in simulation studies (e.g. Hu & Bentler, 1998; Marsh et al., 1988). There are not definite criteria as fit indices are generally affected by the model and data structure as well as the estimation method (Brown, 2015). Notwithstanding that, other indices should be used more as far as possible instead of GFI and AGFI in assessing the goodness of fit. Opting for CFI-IFI, TLI-NNFI, SRMR-RMR duos and RMSEA values will be useful in particular. Besides, considering at least two or more of the suggestions about what fit indices should be used (Brown, 2015; Hu & Bentler, 1999; Kline, 2016) will be an accurate and effective way to evaluate the fit of a model.

As for the values of model fit indices, it was observed that RMSEA, RMR, SRMR, NFI, NNFI-TLI, IFI, and PGFI were reported being at acceptable levels in most of the papers. More than half of the papers in which CFI, GFI, AGFI, and RFI were reported indicated acceptable levels of values regarding the mentioned fit indices. In line with the threshold values referred in the related literature (Hooper et al., 2008; Hu & Bentler, 1999; Kline, 2016; MacCallum et al., 1996; Marsh & Hau, 1996; Steiger, 2007; Tabachnick & Fidell, 2012; Wheaton, 1987), it is possible to conclude that fit indices generally took acceptable values in the papers conducted with CFA,

We make some suggestions here in reference to theory and practice regarding the use of exploratory and confirmatory factor analyses in accordance with the related literature and the results inferred from the papers examined:

- What the analyses assume and whether their data fit to those analyses should be examined. For this purpose, whether some basic assumptions like extreme and missing values, uni- and multi-variate normality, multicollinearity and singularity are assured or not should be checked.
- Methods should be selected considering their appropriateness for data structure. Sample size, continuous or categorical (dual, sequential, etc.) quality of items, and verifiability of basic assumptions should be taken into consideration when opting for a method.
- Adequate sample size should be reached. A sample size of 300 and above is sufficient for both analysis methods, though it may differ according to the number of items. A sample size above 500 is excellent and a sample size of 200 and below is poor and insufficient.

- Analysis results should be interpreted in the light of the related literature.
- In exploratory analysis:
  - Results of KMO ve Bartlet tests, factor eigenvalues, percentages of total variance explained, factor loadings, scree plots, factor extraction and rotation methods used should be reported with construal and rationales of them.
  - More than one methods should be used when determining the number of factors. Choosing the factors whose eigenvalues are greater than 1, carrying out scree plot tests and parallel analysis, incorporating 3 and more acceptable items for a factor are the criteria to be considered.
  - Any items that do not fit to the factor structure should be excluded from the scale one by one and the analyses should be repeated every time. When items are excluded from the analyses, researchers may sequentially observe the items with very low factor loadings (<0.32), items that are factorable independently from the projected factor structure, and items showing high loadings on multiple factors with a difference less than 0.10 between them.
- In confirmatory analysis:
  - The model should be defined accurately.
  - The data structure should be analysed and the assumptions should be controlled.
  - The preferred parameter estimation method should be given with an explanation about why it is used.
  - Factor loadings, track diagram, level of χ^2 significance value and its degree of freedom, the RMSEA value, at least one of the RMR-SRMR values, as well as the CFI and NNFI values should be reported. When evaluating the model, in addition to these minimum criteria, it is also possible to use other indices considering the structure of the model and data.
  - Modifications should not be too much to spoil the psychological structure that is planned to be measured. Reasons should be explained if any modification is to be done.
- If both analysis methods are intended to be used in order to provide stronger evidence for the construct validity, it will be more appropriate to carry out exploratory and confirmatory factor analyses separately over two samples randomly selected from the target population and composed of individuals with the same attributes.

The CFA using papers examined did not provide any evidence about whether the obtained scale structures were proved or not over different groups (gender, socio-economic level, age group, etc.). Therefore, the factor structures obtained over different sub-groups can be hypothetically tested via structural equation modelling. Stronger evidence can be provided for the construct validity of a scale in this way.

Before conducting EFA and CFA, it is of vital importance, in terms of developing more valid and reliable measurement instruments, to examine the assumptions required by those analyses and to have necessary knowledge and equipment regarding how to carry out the analyses and how to interpret their outputs. Researchers should use reliable sources while conducting and interpreting the analyses, and if necessary, they should get support from experts on assessment, evaluation, and psychometry.

Scale development research carried out worldwide in the fields of education, psychology and behavioural sciences and published in certain indices can be examined in further research and their results can be compared with the results in Turkey.

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. doi: 10.17755/esosder.54872

American Psychological Association, American Educational Research Association & National Council on Measurement in Education. (1974). *Standards for educational and psychological tests and manuals.* Washington, DC: American Psychological Association.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*(2), 155-173.

Bailey, K. D. (1982). *Methods of social research* (2nd ed.). New York: The Free Pres.

Bayık, M., & Gürbüz, S. (2016). Ölçek uyarlamada metodoloji sorunu: Yönetim ve örgüt alanında uyarlanan ölçekler üzerinden bir araştırma. *The Journal of Human and Work, 3*(1), 1-20.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Bentler, P. M., & Bonnet, D. C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal 9*(2), 27-40. doi: 10.3316/QRJ0902027.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: The Guilford Press.

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361-379). New York: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136-136.

Büyüköztürk, Ş. (2013). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (18. bs.). Ankara: Pegem Akademi.

Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Researh. *Journal of the Society for Social Work and Research*, *1*(2), 99-103. doi: 10.5243/jsswr.2010.8

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Plenum Press.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.

Çokluk, O., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli SPSS ve LISREL uygulamaları*. Ankara: Pegem Yayıncılık.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

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 ve evaluation*, *10*(7), 1-9.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando: Holt, Rinehart and Winston.

Cronbach, L. J. (1984). *Essentials of psychological testing*. New York: Harper.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Crowley, S. L., & Fan, X. (1997). Structural equation modeling: Basic Concepts and applications in personality assessment research. *Journal of Personality Assessment, 68*(3), 508-31.

Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (1st ed., pp. 621-694). Washington, DC: American Council on Education.

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.

Davidshofer, K. R., & Murphy, C. O. (2005). *Psychological testing: principles and applications*. NJ: Pearson/Prentice-Upper Saddle River.

Delice, A., &Ergene, Ö. (2015). Ölçek geliştirme ve uyarlama çalışmalarının incelenmesi: Matematik eğitimi makaleleri örneği. *Karaelmas Eğitim Bilimleri Dergisi, 3*(1), 60-75.

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi Yayınları.

Erkuş, A. (2016). Ölçek geliştirme ve uyarlama çalışmalarındaki sorunlar ile yazım ve değerlendirilmesi. *Pegem Atıf İndeksi, 0*, 1211-1224. doi: 10.14527/9786053183563.075

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Pscyhological Methods*, *4*(3), 272-299. doi: 10.1037/1082-989X.4.3.272

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286-299.

Fraenkel, J. R., & Wallen, N. E. (2011). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill, Inc.

Gorsuch, R.L. (1983). *Factor analysis* (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates.

Guadagnoli, E., & Velicer, W. F. (1988). Relation to sample size to the stability of component patterns. *Psychological bulletin*, *103*(2), 265-275.

Gül, Ş., & Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim, 40*(178), 85-102.

Hair Jr, J. F., Anderson, R. E., Tatham, R. L., William, C., & Black (1995). *Multivariate data analysis with readings.* New Jersy: Prentice Hall.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424-453.

Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Jöreskog, K. G. (1967). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika 34*(2), 183-202.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109-133.

Kline, R. B. (2016). *Principle and practice of structural equation modelling* (4th ed.). New York, NY: The Guilford Press

MacCallum, R.C., Browne, M.W., & Sugawara, H., M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-49.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611-637. doi: 10.1207/S15327906MBR3604_06

Marsh, H.W., Balla, J.R., ve McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410.

Marsh, H. W., & Hau, K. T. (1996). Assessing goodness of fit: Is parsimony always desirable? *The Journal of Experimental Education*, *64*(4), 364-390.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.

Miles, J., & Shevlin, M. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences, 25*(1), 85-90.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennet, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*(3), 430-45.

Muthén, B., du Toit, S.H.C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* (Technical report). Los Angeles: University of California.

Öztürk, N. B., Eroğlu, M. G., & Kelecioğlu, H. (2015). Eğitim alanında yapılan ölçek uyarlama makalelerinin incelenmesi. *Eğitim ve Bilim, 40*(178), 123-137.

Rummel, R. J. (1988). *Applied factor analysis*. Illinois, Evanston: Northwestern University Press.

Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201-292.

Steiger, J.H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893-98.

Streiner, D. L. (1994). Figuring out factors: the use and misuse of factor analysis. *Canadian Journal of Psychiatry*, *39*(3), 135-140.

Şahin, M., & Boztunç Öztürk, N. (2018). Eğitim alaninda ölçek geliştirme süreci: Bir içerik analizi çalişmasi. K*astamonu Eğitim Dergisi, 26*(1), 191-199.

Tabachnick, B.G., & Fidell, L.S. (2012). *Using multivariate statistics* (6th ed.). Needham Heights, MA: Allyn ve Bacon.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*(2), 197-208.

Ullman, J. B. (2012). Structural equation modeling. In B. G. Tabachnik ve L. S. Fidell (Eds.), *Using multivariate statistics* (6th ed.). Pearson.

Wheaton, B. (1987). Assessment of fit in overidentified models with latent variables. *Sociological Methods ve Research*, *16*(1), 118-154.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *National Institutes of Health*, *76*(6), 913-934. doi: 10.1177/0013164413495237.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806-838. doi: 10.1177/0011000006288127

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, *9*(2), 79-94.

Appendix 1. List of Articles Reviewed in the Study

Ahmetoğlu, E., Mükerrem Ünal, A. ve Ergin, D. Y. (2016). Kaynaştırma uygulamalarının başarısını etkileyen etmenler ölçeğinin geliştirilmesi. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, *6*(2), 167-175.

Akaydın, Ş. ve Kurnaz, H. (2015). Lise öğrencilerine yönelik yazma tutum ölçeği: Geçerlik ve güvenirlik çalışması. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *12*(32), 246-261.

Akbulut, N. ve Aslan, S. (2016). Örtük program ölçeği: Ölçek geliştirme çalışması. *Elektronik Sosyal Bilimler Dergisi*, *15*(56), 169-176.

Akın, A. (2006). 2×2 başarı yönelimleri ölçeği: Geçerlik ve güvenirlik çalışması. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, *12*, 1-13.

Akın, A. (2007). Öz-güven ölçeğinin geliştirilmesi ve psikometrik özellikleri. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *7*(2), 167-176.

Akın, A. (2010). İki boyutlu sosyal istenirlik ölçeğinin geliştirilmesi ve psikometrik özelliklerinin araştırılması. *Gazi Eğitim Fakültesi Dergisi*, *30*(3), 771-784.

Akın, U. ve Yıldırım, N. (2015). Öğretmenlerin eğitime inanma düzeyleri ölçeği (ÖEİDÖ): Geçerlik ve güvenirlik çalışması. *e-International Journal of Educational Research*, *6*(1), 70-83.

Aksoy, V. ve Diken, İ. H. (2009). Rehber öğretmen özel eğitim öz yeterlik ölçeği: Geçerlik ve güvenirlik çalışması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, *10*(1), 29-37.

Altuntaş, S. ve Atli, A. (2015). Eş seçme stratejileri envanteri geliştirilmesi. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *12*(32), 123-135.

Arıcak, O. T., Kınay, H. ve Tanrıkulu, T. (2012). Siber zorbalık ölçeğinin ilk psikometrik bulguları. *Hasan Ali Yücel Eğitim Fakültesi Dergisi*, *17*, 101-114.

Arslan, A. ve Kırık, A. M. (2013). Sosyal paylaşım ağlarında konum belirleme ölçeğinin geçerlik ve güvenirlik çalışması. *Marmara Üniversitesi Öneri Dergisi*, *10*(40), 223-231.

Arslan, G. (2015). Psikolojik istismar ölçeği (PİÖ) geliştirme çalışması: Ergenler için psikometrik özelliklerinin incelenmesi. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, *4*(2), 727. doi: 10.14686/buefad.v4i2.5000146983

Arslan, S. ve Gelişli, Y. (2015). Algılanan öz-düzenleme ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Sakarya University Journal of Education*, *5*(3), 67-74. doi: 10.19126/suje.91303

Atıcı, M. ve Cabaroğlu, N. (2013). Davranış yönetiminde öğretmen yetkinlik algısı ölçeği. *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*, *22*(2), 223-238.

Atli, A. ve Kaya, A. (2016). Lise öğrencilerinin mesleki değerleri. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *12*(1), 314-330. doi: 10.17860/efd.78099

Ayaz, M. F. (2016). Çokkültürlülük algı ölçeği: Geçerlik ve güvenirlik çalışması. *Elektronik Sosyal Bilimler Dergisi*, *15*(57), 463-471. doi: 10.17755/esosder.11595

Bakıoğlu, Ç. ve Kurtuldu, M. K. (2015). Piyano dersine yönelik tutum ölçeği geliştirme çalışması. *Alan Eğitimi Araştırmaları Dergisi (ALEG)*, *1*(1), 33-39.

Balcı Çelik, S., Batık, M. V., Özcan, K. ve Aksoy, S. (2015). Obez çocuklara yönelik tutum ölçeği: Geçerlik ve güvenirlik çalışması. *e-International Journal of Educational Research*, *6*(1), 40-53. http://dergipark.ulakbim.gov.tr/ijer/article/view/1073000378 adresinden erişildi.

Başer, E. H. ve Kılınç, E. (2015). Küresel sosyal sorumluluk ölçeği: Geçerlik ve güvenirlik çalışması. *Sakarya University Journal of Education*, *5*(3), 75-89. doi: 10.19126/suje.57328

Bayoğlu, F. ve Atli, A. (2014). İnönü evlilik tutum ölçeği: Geçerlik ve güvenirlik analizleri. *Ege Eğitim Dergisi*, *15*(2), 397-415.

Bütüner, S. Ö. ve Baki, A. (2011). Development of attitude scale toward the usage of history of mathematics. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *5*(2), 278-311.

Cavkaytar, A., Ardıç, A. ve Aksoy, V. (2014). Aile gereksinimlerinin belirleme aracının geçerlik ve güvenirliğinin güncellenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, *15*(2), 1-12.

Çakıroğlu, Ü., Gökoğlu, S. ve Çebi, A. (2015). Öğretmenlerin teknoloji entegrasyonlarına yönelik temel göstergeler: Bir ölçek geliştirme çalışması. *The Journal of Gazi Education Faculty*, *35*(3), 507-522.

Çankaya, C. ve İşçen, C. F. (2014). Fen bilgisi öğretmen adaylarına yönelik su tüketim davranış ölçeği: Geçerlik ve güvenirlik çalışması. *e-Journal of New World Sciences Academy*, *9*(3), 341-352. doi: 10.12739/NWSA.2014.9.3.1C0622

Çelik, M. ve Yazgan İnanç, B. (2009). Evlilik doyum ölçeği: Geçerlik ve güvenirlik çalışmaları. *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*, *18*(2), 247-269.

Çeliktürk, Z. ve Yamaç, A. (2015). İlkokul ve ortaokul öğrencileri için okuma kaygısı ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Elementary Education Online*, *14*(1), 97-107. doi: 10.17051/io.2015.03320

Çengel, M. ve Türkoğlu, A. (2015). Öğrenciler tarafından algınan sınıf iklimi ölçeği: Geçerlik ve güvenirlik çalışması. *Eğitimde Kuram ve Uygulama*, *11*(4), 1240-1257.

Çetin, B., Bars, M. ve Bars, E. (2015). Matemati□k umutsuzluk ölçeği□ni□n güveni□rli□k ve geçerli□k çalişmasi. *Elektronik Sosyal Bilimler Dergisi*, *14*(53), 163-172.

Çetin, Ş. (2006). Öğretmenlik mesleği tutum ölçeğinin geliştirilmesi (geçerlik ve güvenirlik çalışması). *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, (18), 28-37.

Çıldır Pelitoğlu, F. ve Özgür, S. (2013). İlköğretim öğrencileri için sağlık tutum ölçeği geliştirilmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *7*(1), 32-56.

Çögenli, M. Z. ve Asunakutlu, T. (2014). Akademisyenlere yönelik mobbing ölçeği geçerlik ve güvenirlik çalışması. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, (18), 92-105.

Dağhan, G. ve Akkoyunlu, B. (2016). Çevrimiçi öğrenme ortamlarında kullanım sürekliliğini yordayabilecek yapılara ilişkin geçerlik ve güvenirlik çalışması. *Ege Eğitim Dergisi*, *17*(1), 198-224.

*Dalgıç, G. (2010). Öğretim elemanı performans değerlendirme envanterleri (ÖEPDE) geçerlik ve güvenirlik çalışması. *Elektronik Sosyal Bilimler Dergisi*, *9*(32), 92-105. doi: 10.17755/esosder.68526

Dede, Y. ve Yaman, S. (2008). Fen öğrenmeye yönelik motivasyon ölçeği: Geçerlik ve güvenirlik çalışması. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *2*(1), 19-37.

Demir, N., Kızılay, E. ve Bektaş, O. (2016). Development of an achievement test about solutions for 7th graders: A validity and reliability study. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *10*(1), 209-237. doi: 10.17522/nefefmed.52947

Demirci Seyrek, Ö., Ersanlı, K. ve Tunç, T. (2016). Duyarlık ölçeği. *Elektronik Sosyal Bilimler Dergisi*, *15*(56), 284-294. doi: 10.17755/esosder.41322

Deniş, H., Genç, H. ve Demirkaya, H. (2008). Milli parka yönelik tutum ölçeği geliştirilmesi. *GÜ, Gazi Eğitim Fakültesi Dergisi*, *28*(2), 95-107.

Dilekli, Y. ve Tezci, E. (2015). Öğretmenlerin düşünme becerilerinin öğretimine yönelik sınıf içi uygulamalar ölçeğinin geçerlik ve güvenirlik çalışması. *e-Journal of New World Sciences Academy*, *10*(4), 276-290. doi: 10.12739/NWSA.2015.10.4.1C0647

Doğan, B. ve Çermik, H. (2016). Kitap okumaya yönelik tutum ölçeği: Geçerlik ve güvenirlik çalışması. *e-Journal of New World Sciences Academy*, *11*(3), 168-183. doi: 10.12739/NWSA.2016.11.3.1C0660

Duruhan, K. ve Ersöz, Y. (2013). Ortaokulda demokrasi algıları ölçeği: Geçerlik ve güvenirlik çalışması. *Journal Of History School*, *6*(XVI), 657-672. doi: 10.14225/Joh355

Ekici, E., Taşkın Ekici, F. ve Kara, İ. (2012). Öğretmenlere yönelik bilişim teknolojileri öz-yeterlik algısı ölçeğinin geçerlik ve güvenirlik çalışması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, *31*, 53-65.

Erdem, M. ve Arkün Kocadere, S. (2015). Yapılandırmacı öğrenme inanç ölçeğinin geliştirilmesi. *İlköğretim Online*, *14*(2), 1260-1275. doi: 10.17051/io.2015.84249

Eren, E., Yurtseven Avci, Z. ve Seçkin Kapucu, M. (2014). Developing a scale for competencies and perceptions of necessity about using practical tools for content development. *Journal of Theory & Practice in Education*, *10*(5), 1177-1189.

Erol, M. ve Yıldırım, İ. (2016). The development of higher education life satisfaction scale. *Eğitimde Kuram ve Uygulama*, *12*(1), 221-243.

Ersanlı, K. ve Doğru Çabuker, N. (2015). Diğerkâmlık ölçeğinin psikometrik özellikleri. *Elektronik Sosyal Bilimler Dergisi*, *14*(52), 43-53. http://dergipark.ulakbim.gov.tr/esosder/article/view/5000079930 adresinden erişildi.

Ersanlı, K. ve Uysal, E. (2015). Belirsizliğe karşı tutum ölçeğinin geliştirilmesi. *Elektronik Sosyal Bilimler Dergisi*, *14*(54), 46-53.

Eryılmaz, A. (2015). Amaçlar için mücadele ölçeği. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, *8*(4), 233-244.

Eryilmaz, A. ve Dereli, E. (2011). Psychometric properties of time orientation during classroom disengagement scale. *Elementary Education Online*, *10*(3), 1178-1190.

Eryilmaz, A. ve Mammadov, M. (2016). Türk edebiyatı/Dil ve anlatım dersinde akış durumu ölçeğinin psikometrik özelliklerinin incelenmesi. *21. Yüzyılda Eğitim ve Toplum Eğitim Bilimleri ve Sosyal Araştırmalar Dergisi*, *5*(13), 247-258.

Eryılmaz, A. ve Mammadov, M. (2016). Developing the scale of factors increasing study motivation. *Journal Of European Education*, *6*(2), 73-87. doi: 10.18656/jee.54739

Gençtanırım, D. (2014). Riskli davranışlar ölçeği üniversite formu: Geçerlik ve güvenirlik çalışmaları. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *5*(1), 24-34. http://dergipark.ulakbim.gov.tr/epod/article/view/1040000029 adresinden erişildi.

Girgin, D. (2016). Çalgı eğitiminde algılanan aile desteği ölçeği geliştirme çalışması. *İlköğretim Online*, *15*(3), 778-786. doi: 10.17051/io.2016.90950

Gömleksiz, M. N. ve Kan, A. Ü. (2012). Sosyal bilgiler dersi motivasyon ölçeğinin geçerlik ve güvenirlik çalışması. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, *22*(2), 116-125.

Güner, N., Akkurt Denizli, Z., Sezer, R. ve Bayraktar, A. (2015). 6.- 8. sınıf öğrencilerinin matematik dersi ve matematik ders kitapları hakkındaki görüşlerini belirlemeye yönelik ölçek geliştirme çalışması. *e-Journal of New World Sciences Academy*, *10*(2), 54-65. doi: 10.12739/NWSA.2015.10.2.1C0632

Gürbüz, H., Derman, M. ve Çakmak, M. (2013). Biyoçeşitlilik okuryazarlığı ölçeği: Geliştirme, geçerlik ve güvenirliği. *Elektronik Eğitim Bilimleri Dergisi*, *2*(3), 77-91.

Güven, Y. ve Işık, B. (2006). Beş yaş çocukları için Marmara Sosyal Duygusal Uyum Ölçeğinin (MASDU-5 Yaş) geçerlik ve güvenirlik çalışması. *M. Ü. Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, *23*, 125-142.

*Güvenç, H. (2015). Öğretmen güdüsel desteği ölçeği geliştirme ve uyarlama çalışması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *15*(1), 129-145.

Harman, G. ve Alat, K. (2015). Fen bilgisi öğretmen adaylarının fen ve teknoloji dersinde model kullanımına yönelik tutum ölçeği geliştirilmesi. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, *17*(1), 30. doi: 10.17556/jef.18249

İlgan, A., Aslanargün, E. ve Shaukat, S. (2015). Developing teacher professionalism scale Validation and reliability study. *Eğitimde Kuram ve Uygulama*, *11*(4), 1454-1474.

Kan, A. (2007). Öğretmen adaylarının eğitme-öğretme özyetkinliğine yönelik ölçek geliştirme ve eğitme-öğretme özyetkinlikleri açısından değerlendirilmesi (Mersin Üniversitesi örneği). *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *3*(1), 35-50.

Kapıkıran, Ş. (2016). Ebeveyn akademik başarı baskısı ve desteği ölçeğinin psikometrik değerlendirmeleri ve yapısal geçerlik: Ortaokul ve lise öğrencileri. *Ege Eğitim Dergisi*, *17*(1), 62-83.

Kaplan, A., Öztürk, M. ve Doruk, M. (2014). Matematik uygulamaları dersine yönelik beklenti ölçeği: Geçerlik ve güvenirlik çalışması. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, *16*(2), 85-103. doi: 10.17556/jef.19885

Karaçanta, H. (2009). Öğretmen adayları için kamu personeli seçme sınavı kaygı ölçeğinin geliştirilmesi (geçerlik ve güvenirlik çalışması). *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, (25), 50-57. http://esefdergi.gazi.edu.tr/makaleler/230.pdf adresinden erişildi.

Karaırmak, Ö., Hacıfazlıoğlu, Ö. ve Şirin, S. (2014). Dengeli bütünleşme ve ayrışma ölçeğinin psikometrik özellikleri: Ergenlerde geçerlilik ve güvenirlik çalışması. *Hasan Ali Yücel Eğitim Fakültesi Dergisi*, *11*(22), 237-257.

Karakuş, M. ve Çobanoğlu, N. (2013). İlkokullarda entelektüel sermayenin ölçülmesi ve bazı değişkenler açısından incelenmesi. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, *6*(4), 158-181.

Karaman Kepenekçi, Y. ve Nayir, K. F. (2015). Örgütsel affedicilik ölçeğinin geliştirilmesi ve öğretmenlerin örgütsel affediciliğe ilişkin görüşleri. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *11*(3), 922-934. doi: 10.17860/efd.93627

Karaşar, B. ve Öğülmüş, S. (2016). Sosyal onay ihtiyacı ölçeği: Geçerlik ve güvenirlik analizi. *Ege Eğitim Dergisi*, *17*(1), 84-104.

Karatekin, K. (2013). Öğretmen adayları için katı atık ve geri dönüşüme yönelik tutum ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Uluslararası Avrasya Sosyal Bilimler Dergisi*, *4*(10), 71-90.

Kasapoğlu, F. (2015). Development of spiritual orientation scale: The study of validity and reliability. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *16*(3), 51-68. doi: 10.17679/iuefd.16360640

Kaya, A. ve Bozkur, B. (2015). Kadercilik eğilimi ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *11*(3), 935-946. doi: 10.17860/efd.55137

Kaya, M. ve Doğan, U. (2014). Öğrenci sorumluluk: Ölçek geliştirme, güvenirlik ve geçerlik çalışması. *Journal of European Education*, *4*(1), 11-18.

Keçeci, G., Alan, B. ve Kırbağ Zengin, F. (2016). Eğitsel bilgisayar oyunları destekli kodlama öğrenimine yönelik tutum ölçeği: Geçerlilik ve güvenirlik çalışması. *e-Journal of New World Sciences Academy*, *11*(3), 184-194. doi: 10.12739/NWSA.2016.11.3.1C0661

Keskin, İ., Korkut, A. ve Can, S. (2016). Öğrenme güçlüğü yaşayan öğrencilere yönelik öğretmen yeterlilikleri ölçeğinin geliştirilmesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, *5*(1), 133-155. doi: 10.17539/aej.43261

Koç, F., Sak, R. ve Kayri, M. (2015). Okul öncesi eğitim programındaki etkinliklere yönelik öz-yeterlik ölçeğinin geçerlik ve güvenirlik analizi. *İlköğretim Online*, *14*(2), 1416-1427. doi: 10.17051/io.2015.50571

Kolburan Geçer, A. ve Deveci Topal, A. (2015). E-derslere yönelik memnuniyet ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Eğitimde Kuram ve Uygulama*, *11*(4), 1272-1287.

Korkut Owen, F. ve Bugay, A. (2014). İletişim becerileri ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *10*(2), 51-64. doi: 10.17860/EFD.95021

Kurtuldu, M. K. (2011). Piyano eğitiminde sosyal durum ölçeği geçerlik ve güvenirlik çalışması. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, *31*, 85-96.

Kurtuldu, M. K. (2010). Müzik öğretmeni adayı öz yeterlik ölçeği geçerlik ve güvenirlik çalışması. *e-Journal of New World Sciences Academy*, *5*(2), 178-189.

Küçük, L. ve Miman, M. (2015). Çalışma yaşamında cinsiyetçilik ölçekleri geliştirme ve Toros Üniversitesinde bir uygulama. *Toros Üniversitesi İİSBF Sosyal Bilimler Dergisi*, *2*(4), 117-137.

Melanlıoğlu, D. (2013). Ortaokul öğrencileri için dinleme kaygısı ölçeğinin geçerlik ve güvenirlik çalışması. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *6*(11), 355-361.

Oksal, A., Şenşekerci, E. ve Bilgin, A. (2006). Merkezi epistemolojik inançlar ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, *19*(2), 371-381.

Onur, N. ve Nazik, M. H. (2014). Öğretmenler için bireysel finans alanında finansal tutum ölçeğinin geliştirilmesi (FTÖ): Geçerlik ve güvenilirlik çalışması. *e-Journal of New World Sciences Academy*, *9*(4), 90-99. doi: 10.12739/NWSA.2014.9.4.2C0054

Otrar, M. ve Kandemir, Y. (2016). Özel okullara yönelik bir tutum ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *International Journal of Social Sciences and Education Research*, *2*(1), 284-297.

*Özcan, Z. Ç. ve Erktin, E. (2014). Matematik ödev alışkanlığı ölçeği: Geçerlik ve güvenirlik çalışması. *Boğaziçi Üniversitesi Eğitim Dergisi*, *31*(2), 27-47.

Özer, A., Gençtanırım Kurt, D., Kızıldağ, S., Demirtaş Zorbaz, S., Arıcı Şahin, F., Acar, T. ve Ergene, T. (2016). Ergenler için yeterlik algısı ölçeğinin (YAÖ) geliştirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *7*(1), 237-250. doi: 10.21031/epod.17205

Özeren, E. ve Akpınar, B. (2016). İnternet kullanım alışkanlıkları ölçeği geçerlik ve güvenirlik çalışması. *Elektronik Eğitim Bilimleri Dergisi*, *5*(9), 20-29.

Özlü, G., Özer Keskin, M. ve Gül, A. (2013). Çevre eğitimi öz-yeterlik ölçeği geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, *33*(2), 393-410.

Özsoy-Güneş, Z., Derelioğlu, Y. ve Kırbaşlar, F. G. (2011). İşlemsel fizik ve kimya problemlerinde matematik kullanım ölçeği geliştirilmesi, geçerlik ve güvenirliği çalışması. *Hasan Ali Yücel Eğitim Fakültesi Dergisi*, *16*, 23-38.

Perkmen, S. ve Tezci, E. (2015). Holland teorisinin ışığında meslek kişiliğinin ölçülmesi. *Necatibey Faculty of Education, Electronic Journal of Science and Mathematics Education*, *9*(1), 184-204.

Sarı, M. (2011). Lise yaşam kalitesi ölçeğinin geçerlik ve güvenirlik çalışması. *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*, *20*(3), 253-266.

Sarıçam, H. (2014). Akademik kontrol odağı ölçeği ergen formunun psikometrik özellikleri. *İlköğretim Online*, *13*(4), 1135-1144. doi: 10.17051/io.2014.32429

Seçer, İ. (2015). Zorbalıkla başa çıkma stratejileri ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi*, *30*, 85-105.

Sımsıkı, H. ve Şendil, G. (2014). Baba katılım ölçeğinin (BAKÖ) geliştirilmesi. *Elektronik Sosyal Bilimler Dergisi*, *13*(49), 104-123.

Şahin, F. (2013). The scale for rating the behavioral characteristics of gifted and talented students: study of factor structure, reliability and validity. *Journal of Educational Science*, *38*, 119-132. doi: 10.15285/EBD.2013385570

Şaşmaz Ören, F., Kıyıcı, G., Erdoğmuş, E. ve Sevinç, Ö. S. (2010). Çevre bilincine sahip öğretmen nitelikleri ölçeği: Geçerlik ve güvenirlik çalışması. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *11*(1), 133-152.

Şengül, M. (2013). Ortaokul öğrencilerine yönelik yazma yeterliği ölçeği geliştirme çalışması. *Türkiye Sosyal Araştırmalar Dergisi*, *171*(171), 81-94.

Şimşek-Çetin, Ö. ve Alisinanoğlu, F. (2013). Okul öncesi dönemdeki çocukların yazı farkındalığını değerlendirme kontrol listesinin geçerlik ve güvenirlik Çalışması. *Karadeniz Sosyal Bilimler Dergisi*, *5*(8), 15-27.

Tatar, E., Tosun, C., Tüysüz, C., İlhan, N. ve Karakuyu, Y. (2015). Öğrencilerin fen bilimleri derslerindeki başarılarına etki eden faktörlerin belirlenmesine yönelik bir ölçeğin geliştirilmesi. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, *8*(4), 319-340.

Tataroğlu, B. ve Erduran, A. (2010). Matematik dersinde akıllı tahtaya yönelik tutum ölçeğinin geliştirilmesi. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *1*(3), 233-250. doi: 10.16949/turcomat.65901

Tok, M., Küçük, B. ve Kırmacı, Ö. (2015). Ortaokul kitap okuma alışkanlığı ölçeği: Geçerlik ve güvenirlik çalışması. *Eğitimde Kuram ve Uygulama*, *11*(2), 694-716.

Tösten, R. ve Özgan, H. (2014). Pozitif psikolojik sermaye ölçeği: Geçerlik ve güvenirlik çalışması. *Ekev Akademi Dergi□si□*, *18*(59), 429-442.

Türel, Y. K. (2011). Üniversite öğrencileri içi akıllı tahta kullanımı değerlendirme ölçeği: Geçerlik ve güvenirlik çalışması. *e-Journal of New World Sciences Academy*, *6*(2), 1894-1903.

Türkdoğan, T. ve Duru, E. (2012). Üniversite öğrencileri temel ihtiyaçlar ölçeğinin (ÜÖTİÖ) geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, *31*, 81-91.

Uğurlu, C. T. (2014). İnformal iletişim ölçeği geçerlik ve güvenirlik çalışması. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *15*(3), 83-100. doi: 10.17679/iuefd.64723

Uğurlu, C. T. ve Usta, H. G. (2016). Eğitimde denetim tutum ölçeği geçerlik ve güvenirlik çalışması. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, *18*(1), 137-159. doi: 10.17556/jef.28260

Ünsal, S., Bağçeci, B., Meşe, N. N. ve Korkmaz, F. (2015). Eğitimin işlevlerini değerlendirme ölçeği: Geçerlik ve güvenirlik çalışması. *Adıyaman Üniversitesi Eğitim Bilimleri Dergisi*, *5*(2), 245-265. doi: 10.17984/adyuebd.26853

Varış, Y. A. ve Cesur, D. (2012). Ortaöğretim düzeyi müzik dersine yönelik tutum ölçeği geliştirilmesi. *e-Journal of New World Sciences Academy*, *7*(4), 361-374.

Yalçınalp, S. ve Cabı, E. (2015). Eğitim teknolojileri kullanımı kaygı ölçeği (ETKKÖ): Geçerlik ve güvenirlik çalışması. *İlköğretim Online*, *14*(3), 1005-1016. doi: 10.17051/io.2015.50515

Yalçınkaya, B. ve Eldemir, A. C. (2013). Bireysel çalgı dersine ilişkin tutum ölçeğinin geliştirilmesi. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *10*(21), 29-36.

Yaman, H. ve Dağtaş, A. (2013). Sözlüklere yönelik tutum ölçeği: Geçerlik ve güvenirlik çalışması. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *10*(24), 15-27.

Yaman, H. ve Suroğlu-Sofu, M. (2013). Öğretmen adaylarına yönelik konuşma kaygısı ölçeğinin geliştirilmesi. *Türkiye Sosyal Araştırmalar Dergisi*, *173*(173), 41-50. doi: 10.1080/10967490802095870

Yaman, S. ve Tekin, S. (2010). Öğretmenler için hizmet-içi eğitime yönelik tutum ölçeği geliştirilmesi. *Bayburt Eğitim Fakültesi Dergisi*, *5*(I-II), 76-88.

Yaralı, D. (2015). Özel eğitim dersine ilişkin ölçek geliştirme çalışması: Geçerlik ve güvenirlik. *Kafkas Üniversitesi, e - Kafkas Eğitim Araştırmaları Dergisi*, *2*(2), 12-22.

Yavuz, M., Balkan Kıyıcı, F. ve Atabek Yi□gi□t, E. (2014). İlköğretim II. kademe öğrencileri için çevre okuryazarlığı ölçeği: Ölçek geliştirme ve güvenirlik çalışması. *Sakarya University Journal of Education*, *4*(3), 40-53.

Yeşil, R. (2015). Sivil birey özalgı ölçeğinin geçerlilik ve güvenirlik çalışması. *Journal of Educational Science*, *3*(4), 104-118.

Yeşilyurt, E. (2013). Yapılandırmacı öğrenme kuramına ilişkin bilişsel farkındalık ölçeği geliştirme çalışması: Bir ölçek revizyonu. *e-International Journal of Educational Research*, *8*(2), 285-307. doi: 10.12739/NWSA.2013.8.2.1C058

Yeşilyurt, E. (2012). Yapılandırmacı öğrenme konusunda öğretmen adaylarının yeterliği ölçeği: Geçerlik ve güvenirlik çalışması. *Uluslararası Avrasya Sosyal Bilimler Dergisi*, *3*(7), 29-45.

Yıldırım, B. (2015). Fen bilimleri öğrenme kaygı ölçeği: Geçerlilik ve güvenirlik çalışması. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, *3*(1), 33. doi: 10.18506/anemon.71733

**These studies were dealt with as two studies because more than one scale was developed.*