Item Removal Strategies Conducted in Exploratory Factor Analysis: A Comparative Study

Meltem Acar Guvendir [1,*], **Yesim Ozer Ozkan** [2]

[1]Trakya University, Faculty of Education, Department of Educational Sciences, Turkiye
[2]Gaziantep University, Faculty of Education, Department of Educational Sciences, Turkiye

Abstract: The aim of this study is to examine how the practice of different item removal strategies during exploratory factor analysis (EFA) phase of scale development change the number of factors, factor loadings, explained variance ratio, and reliability values (α and ω) explained. In the study, data obtained from 379 university students were used for the development of a 46-item scale. As the first item removal strategy, crossloading items on two factors and where the difference between factor loadings was less than .10 were identified. Then, items were removed one by one, starting with the item with the least difference between the loadings on the factors. As the second strategy, the items that loaded on two factors and where the difference between factor loadings was less than .10 were found, and these items were removed from the scale as a whole. As the third strategy, the items that gave high loading on more than two factors and where the difference between these factors was less than .10 were identified. The item removal process was started with these items. The study results show that the factor numbers obtained using three different strategies during the item removal process of EFA were the same; however, the number of items on the scale, the explained variance ratio, and the total scale, and reliability values differed. Furthermore, the items in the factors were not all the same. The study results underscore the importance of theoretical competence in the scale development process.

1. INTRODUCTION

Scales are measuring instruments that outline the criteria to be followed when classifying, sorting, or quantifying variables under investigation by researchers. Scales allow for the regulation of the data's quality. According to Netemeyer et al. (2003), while scale development continues to be a popular issue in the social sciences, scholars have proposed a variety of approaches to scale development. Murphy and Davidshofer (2005) divided the scale development process into three stages: instrument construction, instrument standardization, and instrument revision/updating. According to Clark and Watson (1995), scale development is a process that entails clearly defining the target construct, constructing an item pool, testing the pool's items on a representative sample, conducting correlation and factor analysis, and examining the dimension and discrimination validity. Crocker and Algina (1986) state that the scale development process involves determining the purpose of the scale scores, describing the

---

behaviors, preparing the indicator table, creating the item pool, editing the items, administering pre-tests and pilot tests, item analysis, calculating validity and reliability, and standardization steps. Along with the phases outlined by Crocker and Algina (1986), Erkuş (2012) noted that the scale development process should include the following: selection of the scale development technique; written explanations and instructions; and, if required, repeating the applications. Although the process of scale construction is described differently, the fundamental objective is to generate a valid measure of the underlying psychological construct (Clark & Watson, 1995).

The term validity refers to the degree to which a measuring instrument accomplishes its objective. Thus, the measuring instrument is anticipated to measure only the characteristic it is intended to measure. According to Tavsancil (2002), the variables that contribute to the erosion of validity are linked to the development and implementation of scales. Construct validity refers to whether the scale measures precisely what it is intended to measure or whether the items on the scale accurately represent the theoretical or psychological construct (Erkuş et al., 2017). As Messick (1981) phrased it, construct validity is undoubtedly the core of validity. Factor analysis can be used to ascertain the scales' construct validity (Cronbach & Meehl, 1955). The determination of whether a test measures the intended or anticipated construct is a form of validity problem, which may be resolved using factor analytical methods (Stapleton, 1997).

Spearman pioneered the application of factor analysis in the early twentieth century (Ford, MacCallum & Tait, 1986). He claimed that intelligence had a one-factor structure by demonstrating that the individual's varied mental operations had a similar feature. In 1927, Spearman published his own study, "Talents of Man", in which he asserted that intelligence cannot be described by a single factor using the factor analysis technique. Rather than that, it was composed of two factors: general and specific. Following research that refined Spearman's approach concluded that multiple factors should be involved in the functioning of a complex phenomenon like intelligence (Özgüven, 1994). As a result, many researchers began to employ Spearman's approach in the subsequent time, and it developed into a tool for demonstrating the construct validity of data collecting instruments used, particularly in the domains of education and psychology.

Factor analysis is an important statistical operation in the social sciences since it elucidates the quality and validity of measurement. The primary objective of factor analysis is to reduce the number of dimensions (Brown, 2009). As Kerlinger (1979) expressed it, factor analysis is "one of the most powerful methods yet for reducing variable complexity to greater simplicity" (p. 180). Additionally, factor analysis confirms the scale's capacity to assess the construct being measured. Numerous statistical processes such as regression, correlation, discriminant analysis, and difference tests rely on factors extracted from the original data set (Albayrak, 2006). Factor analysis is used to uncover the invisible and immeasurable dimensions concealed underneath the numerous observable and measurable aspects (Johnson & Winchern, 2002). There are two types of factor analysis: exploratory and confirmatory factor analysis. The aim of exploratory factor analysis (EFA) is to identify variables based on their correlations (Kline, 2011; Stewens, 1996; Tabachncik & Fidell, 2001). A previously established hypothesis is tested using confirmatory factor analysis (CFA) based on the correlations between variables (Kline, 2011; Stewens, 1996; Tabachncik & Fidell, 2001). In other words, this method is a procedure for producing a latent variable (factor) using observed variables in a previously built model. It is frequently used in the construction of scales and validity studies, as well as to validate a preconceived construct.

To ensure the measuring tool's validity, each item must measure a single behavior. Therefore, in scale development studies, during the factor creation process using EFA, if an item's level of relationship with more than one factor is larger than the level of relationship with the other

factor, the item should be counted under the factor with the higher level of relationship. According to Tabachnick and Fidell (2001), .32 is a reasonable rule of thumb for the bare minimum loading of a factor item, which corresponds to around 10% crossloading variation with the variance of the other factor items. A "crossloading" item has a loading factor of .32 or more on two or more variables concurrently. When considering whether to exclude a crossloading item from the study, the researcher should consider if there are sufficient strong loaders (.50 or greater) on each component to support the elimination. When there is significant crossloading, it is probable that the items were created insufficiently or that the a priori factor structure was flawed. According to Çokluk et al. (2010), crossloading items are those that have a high loading on more than one component and a difference of less than .10 between these loadings. According to Can (2018), the difference between factor loadings can be regarded as 0.15 if removing the items from the scale does not pose significant problems. To be more precise, these items do not measure a single behavior. As a result, it is critical for construct validity to exclude elements that have an approximate loading on more than one factor (Can, 2016; Kline, 2011; Stewens, 1996; Tabachnick & Fidell, 2001).

The values presented in these studies take a unique approach to item removal. When the examples given by Can (2016), Tabachnick and Fidell (2001), and Büyüköztürk (2007) are examined, the EFA eliminates all crossloading components. In his technical report on factor analysis, Samuels (2017) suggested that the item elimination process should be repeated until no crossloading items remain. Çokluk et al. (2010) discuss in additional detail whether the removal procedure will be carried out sequentially or totally via crossloading. They said that while doing EFA on a single factor, crossloading items must be deleted individually throughout the item extraction procedure. However, there is no definitive rule on whether crossloading elements should be eliminated from the analysis individually or entirely in an EFA for a multidimensional construct. Due to the fact that this condition will vary depending on the measured construct, the researcher will make this conclusion. If a researcher wishes to eliminate an item, it is advantageous to do so from the most crossloading to the least crossloading item (i.e., the item with the smallest difference between the two factor loadings). Raubenheimer (2004) proposed an alternative method for item removal in EFA. In his investigation, he began by removing elements that made a low contribution to the scale's reliability and kept removing items until the reliability value remained constant. He then switched to EFA when the reliability value remained constant. He concluded the procedure by eliminating items that had a high loading on more than one factor in the factor analysis. He then used direct EFA without removing low-reliability items as a second way. When he compared the results of the two methods, he concluded that the scale had a higher reliability value when EFA was performed after the items with low reliability were removed. Additionally, some items retained by the first method were eliminated by the second one. However, the second method required the removal of more items to enhance validity. As a result, researchers do not agree on the appropriate method for removing items. On the other hand, scale development studies in the literature frequently do not specify how to eliminate items during EFA.

Firstly, from which item on the scale should we begin the item removal? As mentioned above, the first strategy is to remove the items one by one, starting from the item that gives a high loading on two factors and where the difference between factor loadings is the smallest. The analysis is repeated after each item is removed from the test. The analysis is completed when no crossloading item is left. As a second strategy, all the crossloading items that give a high load on two factors and where the difference between factor loadings is less than .10 are determined, and these items are excluded from the analysis. Once the crossloading items are removed from the test, the analysis is repeated once again. After the analysis, if there are still crossloading items, they are all removed from the test and the analysis is repeated once more. In addition to these two options, as a third strategy, the items that give a high load on more than

two factors and where the difference between these loadings is less than .10 are excluded from the test. After removing the items that give high loadings on more than two factors, the items are removed from the analysis one by one, starting with the item that gives high loading on two factors, and the difference between the loadings is the least. The analysis is repeated after each item is removed from the analysis. The analysis is terminated when no crossloading item remains. The reason for the item removal process starting with items that impose high loadings on more than two factors is that these items largely weaken the principle that each item should measure a single behavior.

Due to the fact that these three procedures will provide differing results, it is critical to examine the explained variance ratio acquired from each of them. According to Tabachnick and Fidell (2001), the variance ratio is calculated by dividing the sum of a factor's item's factor loading squares by the factor's overall item count. The high variance ratio explained demonstrates the designed scale's factor structure's robustness (Gorsuch, 1983). Scherer, Luther, Wiebe, and Adams (1988) state that the variance ratio in the social sciences should be between 40% and 60%. Büyüköztürk (2007), on the other hand, underlined that the explained variance ratio for unidimensional scales should be 30% or above, but it should be higher for unidimensional scales. As such, it is of importance to determine the explained variance ratio achieved by employing three distinct procedures during scale development research while removing items through EFA. There are a few studies in the literature (e.g. Raubenheimer, 2004) that use a variety of different EFA strategies on the same data set and report comparable results. There is no study in the literature that compares the three aforementioned strategies. Thus, this study will shed light on the literature, particularly on scale development studies, by comparing the factor numbers, loadings, explained variance ratio, and reliability values derived from three distinct item removal procedures used in EFA.

The purpose of this study is to examine how the use of different strategies to remove items during EFA in scale development studies change the number of factors, loadings, explained variance ratio and reliability values (α and ω) explained. For this general purpose of the research, answers to the following questions were sought:

1. What are the factor numbers, loadings, explained variance ratio, communality values, and reliability values obtained when items are removed starting from the most crossloading item to the least crossloading one while conducting EFA in the scale development studies?
2. What are the factor numbers, loadings, explained variance ratio, communality values, and reliability values when all the crossloading items are removed as a whole during EFA in scale development studies?
3. What are the factor numbers, loadings, explained variance ratio, communality values, and reliability values obtained when items are removed, starting from the items that load on more than two factors and where these loadings are close to each other during EFA in the scale development studies?

2. METHOD

2.1. Participants

The study collected data from 379 university students to develop a 46-item scale. The participants are enrolled at a university. While 70.3% of participants are female, 29.7% are male.

2.2. Data Analysis

While conducting EFA, the study compared the results obtained from three different item removal strategies. Firstly, data were tested in terms of EFA's assumptions. These are: the presence of missing data and extreme values, the adequacy of the sample size, and the suitability

of the data to multivariate normality, whether the items are sufficiently correlated, and multicollinearity.

Before proceeding to EFA, it is essential to review the dataset for missing data. There were no missing data in this study's data set. Following that, the study looked for the existence of extreme values. Since the purpose of this study was to determine the number of factors based on the items, Mahalanobis distances were used to find the presence of multivariate extreme values. The Mahalanobis distance measure shows a subject's distance from the centroid, which is obtained using the means of all variables (Tabacknick & Fidell, 1996). By examining Mahalanobis distances, it was determined that 96 subjects could be considered extreme values in this study, and these subjects were eliminated from the analysis.

Numerous viewpoints have been expressed on the appropriate sample size for factor analysis. There are those who argue that the sample size should be determined by reducing the number of items and factors (Bryman & Cramer, 2001; Kline, 1994; Kass & Tinsley, 1979; Nunally, 1978), while others argue that it should be determined using absolute criteria (Comrey & Lee, 1992; Kline, 1994; Tabachnick & Fidell, 2001). According to Kline (1994), a sample size of 200 individuals is often adequate for factor analysis as an absolute criterion, although this number can be dropped to 100 when the number of factors is small yet open. Comrey and Lee (1992) claimed that 50 is a very small sample size, 200 is a medium sample size, 300 is a decent sample size, 500 is a very good sample size, and 1000 is the ideal sample size. Tabachnick and Fidell (2001) state that at least 300 participants are required for EFA. On the other hand, MacCallum et al. (1999) said that sample size is dependent on the characteristics of the data acquired, implying that precise sample size decisions are challenging. According to them, if communalities are strong and each component can be described by four or more items, the sample size can be small; but, if communalities are low, a large sample size will be required. In this case, reaching the largest possible sample is the best way for factor analysis since it cannot be known how high the communalities will be without analysis. The sample size was kept large to ensure that the EFA assumptions were met, and the Kayser Mayer Olkin (KMO) test was employed to assess the sample size's sufficiency. The calculated value was .839. Leech, Barrett, and Morgan (2005) claimed that a KMO value between 0.50-0.60 was insufficient, a value between 0.60-0.70 was poor, a value between 0.70-0.80 was moderate, a value between 0.80-0.90 was good, and a value over 0.90 was exceptional. As a consequence, the sample size in this study was sufficient for EFA based on the KMO value obtained.

The Bartlett's Test of Sphericity examines if the true correlation matrix differs significantly from the unit matrix. If the p value for this test is less than .05, it shows that the matrix of relationships between the items is different from the unit matrix without relations (Can, 2016). Due to the significance of the acquired value (.000), there is a significant difference between the true correlation matrix and the unit matrix in the current investigation.

Additionally, the data should have a multivariate normal distribution. To establish a multivariate normal distribution, the observations in the sample must exhibit a normal distribution across all variable combinations (Çokluk et al., 2010). Mardia's test can be used to check whether the data fit the multivariate normal distribution properly (Mardia, 1970). Due to the significance of Mardia's test result, it was found that the data did not follow the multivariate normal distribution assumption.

Additional measures are available to ascertain whether the items are sufficiently correlated. The first is the anti-image. The anti-image denotes the proportion of variance in an item that is unrelated to another item in the analysis. Obviously, all items should be highly correlated, in order to minimize an item's anti-images. The anti-image matrices address this issue (Sarstedt & Mooi, 2014). Correlations between anti-images are the inverse of partial correlations. In other words, it reflects the pairwise correlation that remains after other variables' influence is

subtracted. A good factor solution is characterized by shared variance/covariance that extends beyond individual pairs of variables to a larger collection of variables. Anti-image correlation matrix diagonals should be greater than 0.5, which is associated with smaller off-diagonal partial pairwise correlations (Can, 2016; Costello & Osborne, 2005; Hauben et al., 2017; Spicer, 2005). In this study, the anti-image correlation matrix's diagonals were all greater than .5 (between .830 and .968). The aggregate data screen indicates that the situation is acceptable.

The data matrix's interrelationships should be sufficient. If the correlation coefficient is less than .33 as a result of an observation to be made in the correlation matrix, no factor analysis will be performed (Can, 2016). Given that the inter-item relationships were found to be between .377 and .684, it was concluded that the number of items with acceptable inter-item relationships was quite high. Additionally, the fact that the matrix's determinant is greater than .0001 indicates the possibility of factor analysis.

Another assumption of EFA is that there should be no problem of variable multicollinearity. To check for multicollinearity, tolerance and VIF values were examined in this direction. Tolerance values were found to range between .241 and .460, and VIF values were found to range between 2.423 and 4.151. Thus, tolerance values greater than .10 (Field, 2005; Mertler & Vannatta, 2005) and VIF values less than 10 (Albayrak, 2005), indicate that multicollinearity is not a concern.

The subsequent step was analysis, as the data set fulfilled th EFA's assumptions. During the EFA process, items were removed in ascending order from the most crossloading to the least crossloading. Then, crossloading items were found and eliminated, followed by a repetition of the analysis. The items were removed starting from the items that gave high load close to each other on more than two factors. The results section includes the factor numbers, loadings, and explained variance ratio derived using the three item elimination procedures. Direct oblimin was used as the factor rotation method in this procedure. Given the fact that oblique rotation creates a pattern matrix including the factor or item loadings, as well as a factor correlation matrix including the factor correlations. Oblique rotation methods such as Direct Oblimin and Promax are prevalent. Direct Oblimin aims to simplify the output's structure and mathematics (Gorsuch, 1983). As a result, the direct oblimin method was selected throughout the factor analysis procedure.

Unweighted Least Squares (ULS) was employed in the factor analysis process. Comrey (1962) developed unweighted least squares analysis in order to minimize the squares of the differences between observed and reproduced correlation matrices. Additionally, it is a subset of principal factor analysis in that it estimates the variance of common factors following analysis (Tabachnick & Fidell, 2001). This technique was used since the data in the research were categorical and violated the multivariate normality assumption.

Additionally, the correlation matrix employed in factor analysis might be vary depending on the number of categories in the scales. As the number of categories rises, the data may be regarded continuous and the Pearson correlation matrix can be used to do analyses. Finney and DiStefano (2013) claim that a data set is deemed continuous if it has six or more categories. Tabachnick and Fidell (2001) state that a data set is considered continuous if it contains seven or more categories. The tetrachoric correlation matrix, on the other hand, is used to analyze data with two categories, whilst the polychoric correlation matrix is used to study data with three, four, or five categories.

Correlation calculated to explain the relationships between unobserved variables is known as polychoric correlation (Basto & Pereira, 2012). Correlations are classified as polychoric if they are based on the premise that ordinal variables have a common continuous distribution (Ekström, 2011). When dealing with ordinal data, factor analysis should be conducted on the

raw data matrix of polychoric correlations, not on Pearson correlations (Basto & Pereira, 2012). Because there were five categories in this study, factor analysis was done using the polychoric correlation matrix (5-point Likert scale-values between 0 and 4).

In factor analysis, when deciding on the number of factors, the eigenvalues-greater-than-one, the communalities explained by the items greater than .2, scatter plot (Cattell, 1966), the minimum average-partial correlation (minimum-average partial correlation), Bartlett's χ2 test (Bartlett, 1950), RMSEA-based maximum-likelihood method (Park et al., 2002), parallel analysis (Horn, 1965), and Velicer's Minimum Average Partial Test (MAP) (Velicer, 1976) can be used. It should be noted, however, that each of these procedures produces a distinct factor number. Additionally, Hayton et al. (2004) state that examining eigenvalues is a commonly used approach for determining the number of factors. However, this technique is insufficient for capturing the true factor structure of the data since it overestimates the number of latent factors. Additionally, the scatter plot is subjective, as it is constructed based on the researcher's observation. As for Bartlett's χ2 test, this test is based on χ2 and is therefore sensitive to sample size. Although the RMSEA based maximum likelihood method is relatively accurate, it can only be used when the maximum likelihood estimator is used. Parallel analysis and the MAP test were used to determine the number of factors in this study, as well as communality and eigenvalues. The factor program and JASP were used for the analysis of this research.

3. RESULT / FINDINGS

To begin, parallel analysis and the MAP test were used to determine the number of factors in the data. While the parallel analysis indicated a two-factor structure, the MAP indicated a three-factor structure. Additionally, the eigenvalues were examined, and it was observed that four factors had eigenvalues greater than one. Table 1 contains the eigenvalues of the four components.

Table 1. *Eigenvalues.*

Variable	Eigenvalue	Proportion of Variance
1	26.386	.574
2	2.791	.061
3	1.881	.041
4	1.351	.029

According to the parallel analysis, there is a structure with two factors, three factors according to the MAP test, and four factors when the eigenvalues are examined. As a result, a three-factor structure was chosen.

In EFA, the item elimination procedure began with the most crossloading item and progressed to the least crossloading item. Table 2 shows the factor loadings for the items obtained prior to the item removal operation. According to Table 2, the 28th item has the highest crossloading, since the difference between the factor loadings is the smallest. As a result, the item removal procedure began with this item. After removing each item, the analysis was repeated. When no crossloading items remained, the analysis was concluded. The obtained factor loadings are summarized in Table 3.

Table 2. *Factor loadings for the items obtained before the item removal process using the first strategy.*

Variable	Factor 1	Factor 2	Factor 3
V1		.682	
V2		.823	
V3		.773	
V4		.745	
V5		.820	
V6		.737	.384
V7		.908	
V8		.853	
V9		.553	
V10		.682	
V11		.832	
V12		.587	
V13		.740	
V14		.755	
V15		.807	
V16		.761	
V17	.461	.384	
V18	.378	.462	
V19	.327	.369	.384
V20	.413	.434	
V21	.411	.467	
V22	.317	.474	
V23	.413	.428	
V24	.369	.411	
V25		.395	.459
V26		.537	
V27	.372	.332	
V28	.451	.465	
V29	.403		.515
V30	.576	.303	
V31	.680		
V32	.653		
V33	.739		
V34	.587		.362
V35	.672		
V36	.733		.419
V37	.820		
V38	.903		
V39	.774		
V40	.661		.413
V41	.843		
V42	.794		
V43	.794		
V44	.697		.361
V45	.815		
V46	.763		

[*]Factor loadings below .30 are not considered (Kline, 1994) and are not shown in the table.
[**] While the item shown in red in the table has a higher loading than one in all three factors, the item shown in green shows the most crossloading item and the items shown in blue show the other crossloading items. In addition, all items shown in blue, red, and green are crossloading. In the first method, the item throwing process began with the item shown in green, and in the third method, it began with the item shown in red. During the item removal process, all blue, red, and green items were discarded in the second method.

Table 3. *Factor loadings for the items obtained using the first strategy.*

Variable	Factor 1	Factor 2
V1	.796	
V2	.889	
V3	.814	
V4	.642	
V5	.881	
V6	.441	
V7	.905	
V8	.841	
V9	.703	
V10	.707	
V11	.740	
V12	.789	
V13	.645	
V14	.704	
V15	.846	
V16	.736	
V18	.594	
V20	.737	
V21	.761	
V22	.537	
V26	.694	
V29		.704
V31	.333	.536
V33	.550	.312
V36		.935
V37		.689
V38		.661
V39		.732
V40		.903
V41		.728
V42		.639
V43		.603
V44		.862
V45	.355	.499

*Factor loadings below .30 are not considered (Kline, 1994) and are not shown in the table. In addition, at least .40 factor loadings (Pett, Lackey, & Sullivan, 2003) were taken as basis for items to be placed under one factor.

According to Table 3, when the item removal process was carried out towards the least crossloading item starting from the most crossloading one, a two-factor structure consisting of 34 items was obtained. Communality values of the items are between .431 and .770. The analysis was repeated 11 times as the analysis was performed after removing each item. The factor loadings of the items are between .441 and .935. The explained variance ratio was found to be 64.844. The explained variance ratio according to the factors is 57.296, 7.547, respectively. There are 22 items in factor 1 and 12 items in factor 2.

Secondly, all crossloading items were identified and all of them were removed as a whole. The item removal procedure was completed by completely deleting crossloading items. The analysis was repeated until there were no remaining crossloading items. Table 4 details the factor loadings of the items.

Table 4. *Factor loadings for the items obtained using the second strategy.*

Variable	Factor 1	Factor 2
V1	.763	
V2	.899	
V3	.786	
V4	.720	
V5	.870	
V6	.622	
V7	.939	
V8	.841	
V9	.625	
V10	.652	
V11	.763	
V12	.682	
V13	.682	
V14	.702	
V15	.797	
V16	.711	
V22	.447	.325
V26	.567	
V30		.603
V31		.713
V32		.634
V33		.589
V35		.629
V37		.875
V38		.921
V39		.862
V40		.732
V41		.906
V42		.826
V43		.810
V45		.721
V46		.627

[*]Factor loadings .30 are not considered (Kline, 1994) and are not shown in the table.

Identifying all crossloading items and removing all of them as a whole was used as the second strategy during the item removal of EFA. Ten crossloading items were detected in the first step, and the study was redone once these items were eliminated. Two crossloading items were excluded from the analysis in the second step. One crossloading item was deleted in the third and fourth rounds. When the analysis was repeated, the results revealed a two-factor construct with 32 items. The items had communality values ranging from.440 to.794. Factor loadings for the items ranged from .447 to .939. The explained variance ratio was determined to be 65.460.

The explained variance ratio for the factors is 58.639 and 6.821, respectively. Factor 1 has 18 items, whereas Factor 2 contains 14 items.

As the last and third item removal strategy, items were deleted starting with those that had a high load on more than two factors and were situated close to one another. The process of item removal began with this crossloading item. Following that, the 25th item was deleted since it loaded on three factors again, and the item removal procedure was finished by deleting each item one by one, starting with the most crossloading item and ending with the least crossloading item. The analysis was repeated nine times until there were no remaining crossloading items. Factor loadings of the items are given in Table 5.

Tablo 5. *Factor loadings for the items obtained using the third strategy.*

Variable	Factor 1	Factor 2
V1	.791	
V2	.901	
V3	.811	
V4	.667	
V5	.879	
V6	.496	
V7	.914	
V8	.836	
V9	.674	
V10	.687	
V11	.741	
V12	.759	
V13	.650	
V14	.693	
V15	.823	
V16	.715	
V18	.534	
V20	.670	
V21	.690	
V22	.504	
V27		.461
V30	.351	.544
V31		.626
V32	.326	.506
V35	.326	.544
V36		.928
V37		.802
V38		.775
V39		.794
V40		.868
V41		.808
V42		.709
V43		.694
V44		.846
V45		.596
V46	.315	.494

[*]Factor loadings values below .30 are not considered (Kline, 1994) and are not shown in the table. In addition, at least .40 factor loadings (Pett, Lackey, & Sullivan, 2003) were taken as basis for items to be placed under one factor.

Finally, the items were removed starting from the items that gave high load close to each other on more than two factors. Subsequently, after the items that gave high loadings close to each other on more than two factors (where the difference between factor loadings was at least .10) were removed from the test, the items with the lowest difference between the factor loadings giving high load on two factors were determined and these items were excluded from the test one by one. As a result, the analysis was repeated nine times and a two-factor structure consisting of 36 items was obtained. The communality values of the items were between .421 and .781. The explained variance ratio was found to be 64.418. The explained variance ratio according to the factors was 57.422 and 6.996. The factor loadings were between .461 and .928. There are 20 items in factor 1 and 16 items in factor 2.

Cronback Alfa and McDonald's Omega reliability values according to the total items and factors with each method of the scale are given in Table 6.

Table 6. *Cronbach alfa and McDonald's omega values.*

	1st method		2nd method		3rd method	
	α	ω	α	ω	α	ω
Factor 1	.962	.962	.952	.952	.958	.959
Factor 2	.943	.944	.953	.953	.956	.957
Total	.971	.971	.971	.971	.973	.973

*Reliability values were written according to the factor order obtained according to the 1st, 2nd and 3rd methods

Table 6 shows that α and ω values are close to each other according to the factors and the scale obtained by using all the three strategies. The α and ω values of the scale obtained in the 3nd strategy are higher than the values obtained in the 1st and 2nd strategies.

4. DISCUSSION and CONCLUSION

EFA is used to reveal the structures of variables that are composed of different components, whose structure is not fully known, but whose existence is also present (Can, 2016). Each item in a measurement tool must measure a single feature. This is an indispensable rule for validity. Therefore, in EFA, crossloading items that give high loading on more than one factor and where the difference between these loadings is less than .10 are removed from the test (Can, 2016; Çokluk et al., 2010; Kline, 2011; Stewens, 1996; Tabachncik & Fidell, 2001). Various procedures are used to identify and eliminate these crossloading elements. Relevant research illustrating the processes of EFA demonstrates that items are removed using a variety of strategies (Büyüköztürk, 2007; Can, 2016; Çokluk et al., 2010; Tabachncik & Fidell, 2001). Accordingly, this research examined how the use of different item removal strategies during EFA in scale development studies changed the number of factors, loadings, variance ratio and reliability values (α and ω) explained.

The study's results suggest that while the factor numbers produced using three distinct procedures throughout the item removal phase of EFA were identical, the scale's item number, explained variance ratio by the factors and the overall scale, and reliability values varied. Additionally, the items in the factors were not identical. For instance, an item that remained in the scale following the second strategy was removed during the first item removal strategy. Additionally, the second strategy had the highest explained variance rate. When reliability values were analyzed, the third strategy produced the greatest values for both Cronbach's alpha and omega coefficients. When the number of items is considered, the third approach yielded the most items. The outcomes of all three strategies exhibit a high degree of resemblance in terms of factor loadings and communality values. The second strategy resulted in the highest factor loading and communality values.

The study's findings emphasize the importance of researchers having a firm grasp of the theoretical framework when determining which item removal approach to use throughout the EFA. In particular, the item selection and removal strategy should be matched to the objective of scale development and to the theoretical conceptualization of the target construct. In their groundbreaking studies on measurement and validity, Cronbach and Meehl (1955) and Loevinger (1957) stated the prominence of theory in measurement and stressed that the latent construct should be grounded in a theoretical framework. A well-grounded theory begins with conceptualizations based on a thorough review of the literature which serves two important purposes (Netemeyer et al., 2003). First, such a review will serve to clarify the nature and range of the content of the target construct. Second, a study of the literature may assist in identifying shortcomings with existing measures and determining whether the suggested scale is indeed necessary (Clark & Watson, 1995). Additional crossloading items can be reviewed and revised if necessary. The findings of the current research show that theoretical knowledge and literature review may serve a third purpose in scale development by influencing researchers' decisions on item removal strategies that they will follow during EFA.

EFA is an analytical statistical method, and in EFA, the process of removing items from the test proceeds mechanically. As a result, if the researcher is theoretically competent about the construct being measured, s/he will also be competent in the item removal process. What the researchers should aim for is that the structure in the theoretical framework overlaps with the data at hand (Tabachnick & Fidell, 2001). Henson and Roberts (2006) stated that very few researchers considered the expected number of factors in a theory. According to them, even though factor analysis is an important process in determining the number of factors in scale development, the theoretically recommended and expected number of factors is also very important. Consequently, the items include pieces of a theory, and EFA is done to check whether the items conform to the theory or not.

Bornstein (1996) suggests that before proceeding to EFA, the researcher should review the items and determine if possessing each item is indeed theoretically relevant. Additionally, the researcher should indicate the degree to which a theoretically significant item adequately describes the conceptual framework. Regarding the item removal procedure in factor analysis, Ziegler (2014) noted that the researcher's theoretical underpinning for the construct being examined comes before the methodologies used during item removal. The initial objective is to establish a theoretical framework for the construct to be examined, and statistical approaches become more significant after the theoretical framework is established. As a consequence, there may be items that the researcher believes are critical in explaining the scale's structure. As a result, experimenting with alternative methods of removing the item may prevent such items from being instantly removed from the test. Even if researchers choose an appropriate approach, they should analyze the items they seek to remove first and then establish whether the item is acceptable for the psychological structure being described. In this instance, researchers with a firm grasp of the theoretical background can select the most appropriate strategy for item removal during EFA.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s). **Ethics Committee Number:** Trakya University, 19.01.2022 - 01/18.

Authorship Contribution Statement

Meltem Acar Guvendir: Methodology, Investigation, Resources, Visualization, Software, Formal Analysis, and Writing -original draft. **Yesim Ozer Ozkan:** Introduction and discussion, Writing -original draft, and Validation.

Orcid

Meltem Acar Guvendir (iD) https://orcid.org/0000-0002-3847-0724

Yeşim Ozer Ozkan (iD) https://orcid.org/0000-0002-7712-658X

REFERENCES

Albayrak, A.S. (2006). *Uygulamalı çok değişkenli istatistik teknikleri* [*Applied multivariate statistical techniques*]. Asil Yayın Dağıtım.

Albayrak, A.S. (2005). Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama [*Alternative to the minimum square technique: a multi-linear connection balanced estimating techniques and an application*]. *Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Dergisi, 1*(1), 105-126.

Bartlett, M.S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, 3,* 77-85.https://doi.org/10.1111/j.2044-8317.1950.tb00285.x

Basto, M., & Pereira, J.M. (2012). An SPSS R-Menu for ordinal factor analysis. *Journal of statistical software, 46*(4), 1-29. https://doi.org/10.18637/jss.v046.i04

Bornstein, R.F. (1996). Face validity in psychological assessment: Implications for a unified model of validity. *American Psychologist, 51*(9), 983-984. https://doi.org/10.1037/0003-066X.51.9.983

Brown, J.D. (2009). Statistics Corner Questions and answers about language testing statistics: Principal components analysis and exploratory factor analysis, In. Definitions, differences, and choices. *Shiken: JALT Testing & Evaluation SIG Newsletter, 13*(1), 19 - 23. https://hosted.jalt.org/test/PDF/Brown30.pdf

Büyüköztürk, Ş. (2007). Veri Analizi El Kitabı [*Data analysis handbook for social sciences*]. Ankara: Pegem Yayınları.

Bryman, A., & Cramer, D. (2011). *Quantitative data analysis with IBM SPSS 17, 18 and 19: A guide for social scientists.* Routledge-Cavendish/Taylor & Francis Group.

Can, A. (2016). *SPSS ile bilimsel araştırma sürecinde nicel veri analizi* [*Quantitative data analysis in the process of scientific research with SPSS*]. Ankara: Pegem Akademi.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Clark, L.A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 187-203). American Psychological Association. https://doi.org/10.1037/14805-012

Comrey, A.L. (1962). The minimum residual method of factor analysis. *Psychological Reports, 11*(1), 15-18. https://doi.org/10.2466/pr0.1962.11.1.15

Comrey, A.L., & Lee, H.B. (1973). *A first course in factor analysis.* Academic Press.

Costello, A.B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation, 10*(7), 1-9. https://doi.org/10.7275/jyj1-4868.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. CBS College Publishing.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik* [*Multivariate statistics for social sciences*]. Pegem Akademi.

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme* [*Measurement and scale development in psychology*]. Pegem Akademi Yayınları.

Erkuş, A., Sünbül, Ö., Sünbül, S.Ö., Yormaz, S., & Aşiret, S. (2017). *Psikolojide ölçme ve ölçek geliştirme II* [*Measurement and scale development in psychology II*]. Pegem Akademi.

Ekström, J. (2011). A Generalized Definition of the Polychoric Correlation Coefficient. *UCLA: Department of Statistics, UCLA*. https://escholarship.org/uc/item/583610fv

Field, A. (2005). *Discovering statistics using SPSS.* (2nd ed.). London: Sage

Finney, S.J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. G. R. Hancock ve R. O. Mueller (Ed.), *Structural equation modeling: A second course* (2nd ed., 439– 492). Charlotte.

Ford, J.K., MacCallum, R.C., & Tait, M. (1986). The Application of exploratory factor analysis in applied psychology: A Critical review and analysis. *Personnel Psychology, 39*(2), 291-314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x

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

Hauben, M., Hung, E., & Hsieh, W.Y. (2017). An exploratory factor analysis of the spontaneous reporting of severe cutaneous adverse reactions. *Therapeutic advances in drug safety*, *8*(1), 4-16. https://doi.org/10.1177/2042098616670799

Hayton, J.C., Allen, D.G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191-205. https://doi.org/10.1177/1094428104263675

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. https://doi.org/10.1177/0013164405282485

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185. https://doi.org/10.1007/BF02289447

Johnson, R.A., & Wichern, D.W. (2002). *Applied multivariate statistical analysis.* Upper Saddle River.

Kass, R.A., & Tinsley, H.E.A. (1979). Factor analysis. *Journal of Leisure Research, 11,* 120-138. https://doi.org/10.1080/00222216.1979.11969385

Kerlinger, F.N. (1979). *Behavioral research: A conceptual approach.* Rinehart & Winston.

Kline, P. (1994). *An easy guide to factor analysis.* Routledge.

Kline, R. B. (2011). *Principles and practice of structural equation modeling (3rd Edition).* The Guilford Press.

Leech, N.L., Barrett, K.C., & Morgan, G.A. (2005) *SPSS for Intermediate Statistics, Use and Interpretation. 2nd Edition*. Lawrence Erlbaum.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(3), 635–694. https://doi.org/10.2466/pr0.1957.3.3.635

MacCallum, R.C., Widaman, K.F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1)*,* 84-99. https://doi.org/10.1037/1082-989X.4.1.84

Mardia, K.V. (1970). Measures of multivariate skewnees and kurtosis with applications. *Biometrika, 57*(3), 519-530. https://doi.org/10.2307/2334770

Mertler, C.A., & Vannatta, R.A. (2005). *Advanced and multivariate statistical methods: Practical application and interpretation (3th ed.).* Pyrczak Publishing.

Messick, S. (1981). Evidence and ethics in the evaluation of tests 1. *ETS Research Report Series, 1981*(1), 1-41. https://doi.org/10.1002/j.2333-8504.1981.tb01244.x

Murphy, K.R., & Davidshofer, C.O. (2005). *Psychological testing: principles and applications.* Pearson Education International.

Netemeyer, R.G., Bearden, W.O., & Sharma, S. (2003). *Scaling procedures.* SAGE Publications, Inc.

Nunally, J.C. (1978). *Psychometric theory.* McGraw Hill.

Özgüven, E. (1994). *Psikolojik testler* [*Psychological tests*]*.* Yeni Doğuş Matbaası.

Park, H.S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research, 28*(4), 562-577. https://doi.org/10.1111/j.1468-2958.2002.tb00824.x

Pett, M.A., Lackey, N.R., & Sullivan, J.J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. SAGE.

Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology, 30*(4), 59-64. https://doi.org/10.4102/saji p.v30i4.168

Samuels, P. (2017). *Advice on Exploratory Factor Analysis. Technical Report.* Centre for Academic Success, Birmingham City University.

Scherer, R.F., Luther, D.C., Wiebe, F.A., & Adams, J.S. (1988). Dimensionality of coping: Factor stability using the ways of coping questionnaire. *Psychological Reports, 62*(3), 763-770. https://doi.org/10.2466/pr0.1988.62.3.763

Sarstedt M., & Mooi E. (2014). Factor analysis. *In: A concise guide to market research. springer texts in business and economics*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-53965-7_8

Spicer J. (2005). *Making sense of multivariate data analysis: An Intuitive approach*. SAGE.

Stapleton, C.D. (1997). *Basic concepts and procedures of confirmatory factor analysis* [*Paper presentation*]. The Annual Meeting of the South West Educational Research Association. Austin.

Stewens, J. (1996). *Appied multivariate statistics for the social science (Third Edition).* Lawrence Erlbaum Associates.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlilik* [*Reliability and validity in social and behavioral assessments*]. Seçkin Yayıncılık.

Tabachnick, B.G., & Fidell, L.S. (1996). *Using multivariate statistics*. Harper Collins.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics. 4th Edition, Allyn and Bacon.* MA.

Tavşancıl E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi* [*Measurement of attitudes and data analysis with SPSS*]. Nobel Yayınevi.

Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*(3), 321-327. https://doi.org/10.1007/BF02293557

Ziegler, M. (2014). Comments on item selection procedures. *European Journal of Psychological Assessment, 30*(1), 1-2. https://doi.org/10.1027/1015-5759/a000196