The use of polychoric and Pearson correlation matrices in the determination of construct validity of Likert type scales

Hasan Fehmi Özdemir [iD]
Ankara University, Ankara, Turkey, hozdemir@ankara.edu.tr
Çetin Toraman [iD]
Çanakkale Onsekiz Mart University, Çanakkale, Turkey, toramanacademic@gmail.com
Ömer Kutlu [iD]
Ankara University, Ankara, Turkey, omer.kutlu@ankara.edu.tr

Check for Updates

ABSTRACT   No matter how strong the theoretical infrastructure of a study is, if the measurement instruments do not have the necessary psychometric qualities, there will be a question of trust in interpreting the findings, and it will be inevitable to make wrong decisions with the results. One of the important steps in scale development/adaptation studies is to provide evidence of the experimental validity. In order to reveal evidence of construct validity of Likert type scales, to identify factor structures, to confirm previously predicted structures, factor analysis is used. The primary issue to be examined is the level of measurement of the variable and one of the leading decisions that must be taken is which relation matrix will be used. This descriptive research is based on the effects of using Pearson or polychoric correlation matrix in the factor analysis. It is determined that items showed different "item-total correlations", "loading values" and "correlation coefficients", different factor numbers emerged, different items were removed out of the scale, confirmation status of the structure has changed.

Keywords: Factor analysis, Polychoric correlation matrix, Pearson correlation matrix, Likert type scales, Construct validity

Likert tipi ölçeklerin yapı geçerliğinin belirlenmesi sürecinde polikorik ve Pearson korelasyon matrisinin kullanımı

ÖZ   Bir bilimsel çalışmanın teorik altyapısı ne kadar sağlam olursa olsun kullanılan ölçme araçları gerekli psikometrik nitelikleri taşımıyorsa, bulguların yorumlanmasında güven problemi olacak, elde edilen sonuçlarla yanlış kararlar alınması ise kaçınılmaz olacaktır. Ölçek geliştirme-uyarlama çalışmalarında, kuşkusuz en önemli adımlardan biri, aracın psikometrik niteliklerine dair deneysel geçerlilik kanıtlarının ortaya konmasıdır. Bu niteliklerden biri de yapı geçerliğidir. Likert tipi ölçek geliştirme-uyarlama çalışmalarında, yapı geçerliğine ilişkin kanıtların ortaya konması, faktör yapılarının ortaya çıkarılması ya da daha önceden kestirilen faktör yapılarının doğrulanması amacıyla faktör analizi kullanılır. Faktör analizi öncesinde sorgulanması gereken hususların başında verilerin hangi ölçek düzeyinde toplandığı gelmektedir. Analiz sürecinde alınması gereken önemli kararlardan biri ise hangi ilişki matrisinin kullanılacağıdır. Faktör analizinde, Pearson ya da polikorik korelasyon matrisi kullanmanın analiz sonuçları üzerindeki etkisini incelemeyi ve sonuçlarını karşılaştırmayı temel alan bu araştırma betimsel bir araştırmadır. Farklı korelasyon matrisi temelli faktör analizi sonuçlarının birbirinden farklılaştığı, maddelerin farklı "madde toplam korelasyonu", "yük değeri" ve farklı yönde korelasyon değeri gösterebildiği, farklı faktör sayılarının ortaya çıktığı, farklı maddelerin ölçek dışında bırakılabildiği ve test edilen yapının doğrulanma durumunun değiştiği belirlenmiştir.

Anahtar   Faktör analizi, Polikorik korelasyon matrisi, Pearson korelasyon matrisi, Likert tipi ölçekler, Yapı
Kelimeler: geçerliği

180

INTRODUCTION

Psychological measurement instruments are used in many fields of social sciences, which study human behaviors, especially in psychology, educational sciences and sociology, and also in psychiatry, which has a close relationship with social sciences, within its biopsychosocial model. In general terms, psychological tests are systematic approaches that provides information about individuals' abilities, skills, performances, motivations and attitudes, and contribute to make various decisions according to what is obtained from them (Öner, 1987). Studies regarding scales have an important place in every area where psychological activities are conducted.

There are numerous measurement instruments, which have been developed or adapted for different purposes. In addition to the existing ones, there is need to new instruments to be used in the new studies, which especially assess different psychological features, and which can be used for different samples and age groups etc. This need is met through the development of new measurement instruments or adaptation of the existing ones. While both methods have advantages and disadvantages, the most important thing is that validity, reliability and even possible standardization evidences should be established for these instruments by use of appropriate methods and following the correct processes. No matter how strong the theoretical infrastructure of a scientific study is, if the measurement instruments that are used to collect data do not have the necessary psychometric qualities, there will be a question of trust in interpreting the findings obtained by that study, and it will be inevitable to make wrong decisions with the results obtained from the application of these instruments.

Undoubtedly, one of the most important steps in scale development or adaptation studies is to provide evidence of experimental validity of the psychometric properties of the developed or adapted instrument (Crocker & Algina, 2008). In this context, it is necessary to examine whether the instrument measures the intended feature exactly and accurately. The determination of other psychometric qualities of a scale can be done after this examination. Healthy and precise decisions about the scores to be drawn from the application of an instrument can only be taken in the light of the validity evidences.

The notion of validity, which began to evolve in the 1930s thanks to the efforts of practitioners who wanted to break hegemony of the academics in the American Psychological Association (APA), was first defined in 1954 in testing standards report published by a commission established by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (AEAB, APB, EÖUK, 1997). Although the scope is defined under four categories, namely the content, predictive, concurrent and construct validity, some researchers argue that the concept of validity must be gathered under the framework of construct validity through ongoing discussions (Cronbach & Meehl, 1955; Cronbach, 1984; Messick, 1988, 1993; Şencan, 2005; Urbina, 2014). The investigation process of construct validity is not much different from following the general scientific procedures for development and/or confirmation of a theory (Lemke and Wiersma, 1976). In addition, all the information provided by any validity procedure is relevant to construct validity (Anastasi, 1988). This gives construct validity both a comprehensive status, and also importance and reliability. Therefore, the rest of the validity types can be subsumed under the concept of construct validity. Construct validity has also been associated with factor analysis, which is the most commonly used method to obtain evidence of such validity by many researchers who have even used the concept of factor validity instead of construct validity (Nunnally and Bernstein, 1994).

In scale development and adaptation studies in behavioral sciences; Factor Analysis (FA), which is one of the multivariate statistical methods that Nunnally and Bernstein (1994, p. 111) defined as the heart of the measurement of psychological constructs to reveal evidence of construct validity, to identify factor structures, or to confirm previously predicted factor structures, is used. FA is used in a large number of scientific studies dealing with complex research questions along with other multivariate statistics.

181

Naturally, in the social sciences such as education, psychology and sociology, where multiple variables are being studied simultaneously and the univariate statistical processes are insufficient to explain; multivariate methods such as multivariate analysis of variance, regression analysis, discriminant analysis, multidimensional scaling and factor analysis are needed. Assuming that these methods together constitute a conceptual basis of the topics covered in scientific research and an arch-like structure that leads to the understanding of causality principles, it can be said that the keystone of this structure is the "factor analysis" (Cattell, 1978).

The purpose of this research is to compare the effect of using Pearson or polychoric correlation matrix on the results of the factor analysis, which is applied in order to provide evidence of construct validity during the development process of likert type scales used in social sciences. Although, similar studies aiming to illustrate the advantages of using polychoric rather than Pearson correlations by means of simulation studies held in the field (Holgado-Tello et al., 2010), one of the most important points that differentiates this study from the previous ones is working with real data. Also some new issues in the scope of this study, such as number of items remained and items' additivity states, which were not analyzed in previous simulation studies, increases the importance of this work.

Factor Analysis

In data reduction with FA, unlike many statistical methods that are used to examine the relationship between dependent and independent variables, there is a summarizing process that aims to understand the underlying cause-and-effect relationships in data sets. Gorsuch (1974) stated that the main purpose of the FA is to develop both the theoretical constructs in a particular field and to reveal the operational representatives of these theoretical constructs. Cattell's (1978, p. 4) definition of FA as the queen of correlational methods is supportive of Pedahazur and Schmelkin's (1991) argument, which states that one of the best method to be used in examining the internal structure of a group of variables or indicators is the factor analytical method. It can be argued that the FA is a statistical technique used to demonstrate construct validity (Atılgan, Kan, & Doğan, 2006; Crocker & Algina, 2008; Bowden, 2004; Erkuş, 2003; Kieffer, 1999; Pedhazur & Pedhazur Schmelkin, 1991; Urbina, 2014). The information obtained in the FA provides a road map for subsequent validity and reliability studies and other statistical analyses to be performed on the basis of the scores obtained from the measurement instrument (Çokluk, Şekercioğlu, & Büyüköztürk, 2010, p.177).

Factor analysis, one of the multivariate statistical methods with a wide range of applications, was developed for use in the field of psychology with the groundbreaking studies of Pearson (1901) and Spearman (1904) at the very beginning of the 20th century. However, it has been used in various scientific fields since the second half of the 20th century. Factor analysis is a set of methods rather than a specific scientific method for determining the fundamental dimensions of a data matrix structure. In particular, researchers, who need to explain an individual's behavior, intelligence, and abilities in a mathematical model, have been compelled to develop this scientific method (Albayrak, 2006, p. 107).

Factor analysis can also be defined as a multivariate statistic that aims to find and discover a smaller number of conceptually meaningful new variable(s) by combining a large number of interrelated variables (Büyüköztürk, 2002). It is divided into two general categories: "exploratory factor analysis" and "confirmatory factor analysis". The exploratory factor analysis attempts to discover the connection between observed variables with unknown latent variables, while the confirmatory factor analysis attempts to confirm the aforementioned structure with the data obtained from the measurement instruments (Çokluk, Şekercioğlu, & Büyüköztürk, 2010). In addition to these two types of factor analysis, in modern factor analysis we see hybrid factor analysis, in which confirmatory rotation methods are used after extraction method (as cited in Henson & Roberts, 2006, p. 395).

182

Exploratory Factor Analysis (EFA)

When solving the encountered problems, humankind try to reach the solution by reducing the total work. Known as "the Least Effort Law" in the scientific world, we find this theory in all scientific fields as an attempt to explain the most by using the least. Behavioral scientists also try to explain the observed variables with fewer latent variables. In doing so, they use exploratory factor analysis to derive new and independent k variables (factor), equal or less, by taking advantage of the correlation or covariance matrix attained from data set which consists of *j* associated variables (Özdamar, 2013). Since latent structures or factors are thought to summarize the observed variables, they are closely related to the evaluation of the validity of the theory development and observed scores. Theorizing and measurement of structures are processes linked to each other through organic bonds (Henson & Roberts, 2006). As pointed out by Kieffer (1999), the use of factor analytical techniques in social sciences has been integrated with the evaluation of construct validity of both theory development and measurement. Whatever the purpose of the use of exploratory factor analysis is, the significance of the latent variables is directly related to researchers' definition. As Mulaik (1987) states, it is not the EFA that defines things about psychological features such as intelligence, personality, but it is actually the researchers themselves who make definitions for taking decisions about how to use such concepts. The analytical results, as Thompson and Daniel (1996) argue, will provide information for the definitions to be made, but nevertheless the researcher has full responsibility for the decisions made in this elaboration process. At this point, the role of exploratory factor analysis does not go beyond being a tool. For these reasons, researchers' decisions are needed to be thought thoroughly in the process of exploratory factor analysis (Henson & Roberts, 2006).

Despite having a fairly wide range of uses, the use of exploratory factor analysis in research is under serious criticism (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Most of these criticisms focus on the subjective decision-making necessity arising from the nature of analysis during the execution of exploratory factor analysis. Tabachnick and Fidell (2013) described the absence of a criterion variable that could be used to test outcomes as one of the problems experienced in factor analysis. The interpretation of the analysis results is, to a large extent, based on the presumed judgments of the researchers, who are also assumed to know the analysis sufficiently.

Researchers, who will apply exploratory factor analysis in their studies, should question some basic concepts, follow certain steps in the analysis process, and make decisions throughout the process. In addition to these, they need to be able to master accurate and complete reporting practices in order to increase the contributions they will make to the literature and to the subsequent studies.

At the beginning of the analysis, the primary issues to be examined are the level of measurement of the data, sampling size, missing values and/or outliers, normality, linearity, multi-collinearity and singularity. How to determine the sample size, and whether it is sufficient or not; how to follow a path if there are missing values and/or outliers; how to test the normality of the data, how to obtain useful results from the data set in cases where the normal distribution cannot be obtained, and how to differentiate conditions where the multivariate normal distribution is a necessity are the issues researchers consider.

In the process of analysis, the leading decisions that must be taken by the researchers are; which relation matrix (which correlation matrix or variance-covariance matrix) will be used, which factors in predicting factor/factorization method to be used (main/principal component method, basic/common factor analysis, maximum likelihood method, generalized least squares method, principal axis factorization method, alpha factorization method, image factorization method), which rules will be based on when determining the number of factors (Kaiser criterion/rule-all factors with eigenvalues greater than one, Cattell Scree test or scree plot, explained variance criterion, Joliffe's criterion), in which cases the factor rotation to be applied, while doing this which rotation method (orthogonal or oblique) will be selected (varimax, quartimax and equamax or direct obliques and promax), and the theoretical and practical basis

for the decisions that might largely differentiate results, assumptions and the advantages and disadvantages should also be known.

Confirmatory Factor Analysis (CFA)

In contrast to EFA, which produces theory, CFA is another factor analysis method that tests the theory. The factor analysis method, which is used to reach the answers of the questions, such as whether the relation between the factors belonging to a structure revealed by EFA and the variables are sufficient or not, which variables are related to which factors, whether the factors are independent of each other, whether these factors are sufficient enough to explain the original structure, is called CFA (Özdamar, 2013).

Exploratory factor analysis is used to reveal the best factor model for the observed data set when researchers do not have any idea of the underlying factors before the application of a psychometric measurement instrument. On the other hand, in the event that there is a previously defined and bounded structure, in other words a hypothesis about the underlying factors, then the methodology that the researchers refer to when testing this model/factor structure systematically is confirmatory factor analysis (Bryant, Yarnold, & Michelson, 1999). CFA is used to test a previously validated and reliable instrument's usability in a new culture, in a field, and/or to a target group. It has been indicated that CFA is a more appropriate method to be used by researchers in assessing the construct validity (Stapleton, 1997), and it has been emphasized that between the two main factor analytical methods, CFA is both theoretically more important and should be used more widely (Gorsuch, 1983).

CFA offers important advantages such as comparing the different factorial structures put forth by EFA, providing an opportunity to correct the conceptual and statistical susceptibility of the model, which has been thought to be weak, by transforming it into an improved model with a more reliable and efficient structure.

As in the EFA, it is necessary for the researchers to be in control of a number of issues to be done before and after the analysis in the CFA, and also accurate and complete reporting practices in order to increase the contributions they will provide to the literature and for the follow-up studies.

Before CFA, the relationship matrix to be used depending on the level of measurement of the data available should be determined. The analysis then begins with a description of the model in the direction of theoretical bases. In doing so, the model is determined by fixing or releasing certain parameters (factor coefficients, factor correlation coefficients, variance-covariance of the measurement error) in line with the theoretical expectations of a researcher. This is followed by an analysis of the fit statistics obtained from the estimations of the collected data set and model parameters by using certain programs (AMOS, LISREL, etc.). By using different fit statistics (Chi-Square Goodness of Fit, Goodness-of-Fit Index-GFI, the Root Mean Square Error of Approximation-RMSEA, Residual Averages and Root Mean Square Residual-RMR, Comparative Fit Index-CFI, Normalized Compliance Index-NFI and Non-Normative Compliance Index-NNFI and Parsimony Goodness-of-Fit Index-PGFI), the analysis is continued by evaluating the resulting model fit. If the fit indices cannot meet the acceptance levels, the modification indices must first be evaluated and, if necessary, the model must be redefined and the process must be repeated from the first step. Factor structure, the number of factors, the degree to which each item has a high load to which factor, the model supported by a previous study or a developed theory, and even the amount of error should be implicitly defined.

As in all other statistical methods, the leading issue to be questioned before the factor analysis is at what measurement level the data are collected. Theoretical bases and methods used for factor analysis with continuous variables have been fairly developed. In practice, however, some of the observed and/or measured variables are at the level of ordinal scale, which is often overlooked and is incorrectly treated as if they had the numerical metric property representing the sequential categories like 1, 2, 3, 4 (Jöreskog & Moustaki, 2001). The use of appropriate statistical techniques by accepting the data

184

obtained from these scales as if they are at interval levels, especially due to the item response formats of the scales (Likert type, etc.), which are used frequently in all fields covered by social sciences has been criticized in advance. Researchers who have expressed this criticism indicate that the data obtained from such scales are at the ordinal level of measurement and that appropriate statistical techniques should be used for these data (Stevens, 1946; Thomas, 1982; Jamieson, 2004). When the ordinal scale level variables considered, the necessity of using "tetrachoric" (for two categorical data) or "polychoric" (for three or more categorical data) correlation matrices have been emphasized, while estimating the relationship between variables or conducting correlation-based analysis (Holgado-Tello, Chacón-Moscoso, Barbero- García, & Vila-Abad, 2010; Uebarsax, 2015). Jöreskog and Sörbom (2002) found that the most consistent, reliable, and strongest predictors of factor analysis can be achieved by using the polychoric correlation matrix. However, the Pearson correlation matrix is often used during the application of both exploratory and confirmatory factor analysis used in the scale development and adaptation processes. When the level of the relationship between categorical data are studied, various reasons are suggested why Pearson correlations are not appropriate (Holgado-Tello et al., 2010). Firstly, categorical variables are variables providing information at the level of ordinal scale, and Pearson correlation requires measurements at equal interval level. The information obtained from categorical variables is limited only to the number of observations per cell/categorization in the contingency table. If Pearson correlation is used under these conditions, the restriction/limitation of categorization/classification will lead to the artificial restriction/limitation of the relationship between measurements (Guilley & Uhlig, 1993), all observations / persons placed on the intervals, where the boundaries of each category determined, will be scored by accepting in one of the categories, which will result in a reduction in variability, in other words, a loss of data, even if the data are actually different from each other. Considering that in homogeneous samples, Pearson's correlation gets lower values than usual, the restrictions/limitations arising from this way of assigning scores to the observations/individuals will cause to determine the degree of the relationship between the observed variables lower as well, and consequently it will cause a decline/decrease in the factor loadings obtained by factorization of the correlation matrix (DiStefano, 2002). For this reason, when factor analysis will be used to test the validity of an instrument, it is of paramount importance that the measurement level at which the data is collected is taken into account. On the other hand, frequent use of the Pearson correlation matrix in analyses is due to the fact that researchers do not have enough knowledge about the subject or the limitations of the computer package programs they use. The best example of this is the fact that IBM-SPSS (Statistical Package for the Social Sciences), one of the most popular statistical package programs applied in the social sciences, does not recognize a preference for the correlation matrix to be used during factor analysis, and uses Pearson correlation matrix as a default. Significant amount of measurement error, especially random and systematic errors, are intermingled from different sources, in social sciences, and distorting the estimates of the relationship between the variables involved in the research, which leads to incorrect results. In this context, at the beginning of the negative consequences of the use of different types of correlations in the statistical analysis in terms of methodological investigations, it is possible that the researchers make mistaken inferences about the construct validity, which is the cornerstone of the basic and applied sciences.

METHODOLOGY

This study is a descriptive research. Descriptive research is the study of describing existing conditions without being interested in the relationships or differences between variables. In this context, descriptive studies serve for the purpose of describing science, as well as to provide an insight into the production of hypothesis for further researches (Erkuş, 2013).

Participants

A methodological comparison is planned in this research. The purpose of this research is neither to adapt a new instrument from a different culture, nor to develop a new instrument. For this reason, it has been deemed appropriate to carry out the methodological discussion through an instrument, which has been already developed. The data obtained during the development process of the Gender Equality Scale (GES), which was developed by Gözütok, Toraman and Acar Erdol (2017), were also used in this research with the permission of the researchers. Therefore, the participants are the same participants whose data were collected for the purpose of conducting exploratory and confirmatory factor analyzes in the mentioned study.

In Gözütok et al. (2017) study, data were collected from two separate groups of high school students to perform exploratory and confirmatory factor analysis. The group to collect data for the exploratory factor analysis consists of 396 students and for confirmatory factor analysis consists of 265 students. It was also noted that while groups were formed, groups were heterogeneous (different genders, different class levels [9th, 10th, 11th and 12th grade], in terms of cultural activities such as going to cinema, theater, reading newspapers and books) and balanced group distribution in terms of gender.

Data Collection Tools

In this study, the Gender Equality Scale (GES), and the data collected for factor analysis during the development of this scale were used.

Analysis of the Data

Gözütok et al. (2017) study was analyzed through IBM-SPSS based on the Pearson correlation matrix. Typical analyses used in many scale development studies were used in this analysis. They are namely; examination of item total correlations, Kaiser-Meyer-Olkin (KMO), Bartlett Sphericity test, examination of eigenvalues, varimax axis rotation. The Cronbach Alpha reliability coefficient was used to obtain evidence of reliability (Büyüköztürk, 2013; Özdamar, 2013). Factor analysis findings and results of Gözütok et al. (2017) were used for comparison with the permission of researchers.

In the factor analysis, "FACTOR" software developed by Lorenzo-Seva and Ferrando (2006) was used to obtain a comparative polychoric correlation matrix and to carry out the analyses. This software is free, useful and small-only focuses on factor analysis. By means of this software, it is possible to obtain a tetrachoric correlation matrix so that factor analysis of achievement tests, which can be coded as 0 and 1, can be realized. In addition, as in the case of Likert scales, which have multiple categories, a polychoric correlation matrix can be obtained and factor analysis is carried out over this matrix. The data file was uploaded onto the software and analyzed. In the analysis, the total correlations of the items were examined, the eigenvalues were examined, the Kaiser-Meyer-Olkin (KMO) and Bartlett Sphericity tests were performed, and the varimax rotation method was used when multiple factors were tested. The reliability of the factor software and the model fit indices were taken into account.

The data collected for the CFA were transferred to the AMOS 22 program, and the analyses were carried out with this program.

FINDINGS

According to the researchers who argue that when the measurement instruments developed in social sciences are at the ordinal scale, it is necessary to use the polychoric correlation matrix in the development process of Likert scales (Holgado-Tello, Chacón-Moscoso, Barbero- García, & Vila-Abad, 2010; Jöreskog & Sörbom, 2002; Uebarsax, 2015). Based on this argument, the first analysis was conducted without the number of factors being specified. The FACTOR software not only proposes factor numbers, provides information on eigenvalues, but also gives information regarding the contribution of the items (communality) to scale. As shown in Figure 1, after reviewing this information, the number of factors and the items to be deducted from the analysis were decided.
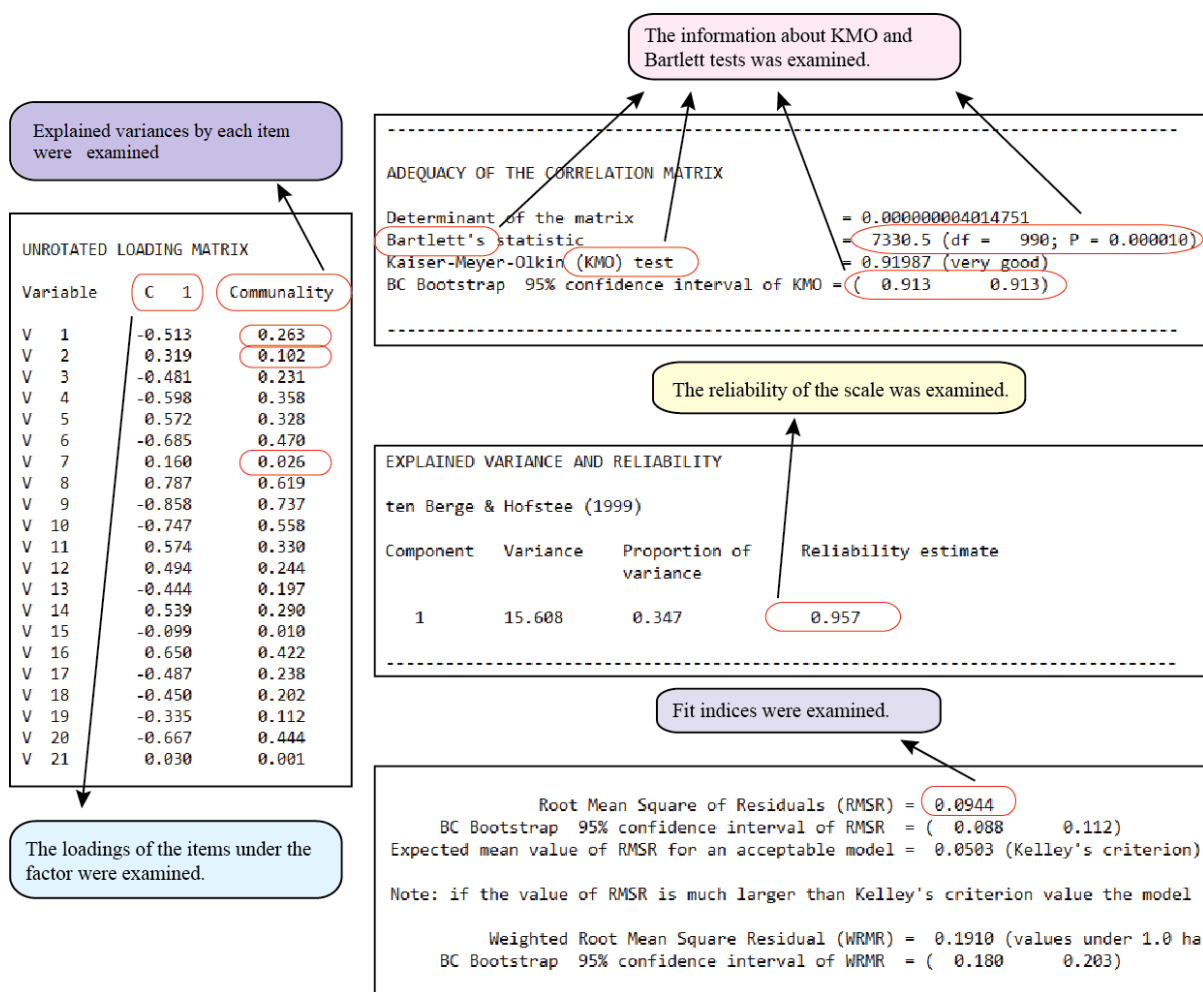
Figure 1. Analysis output file of "FACTOR" software

At the end of the contribution evaluation, where the input values and the items are scaled without specifying the factor number; it is understood that items 1, 2, 3, 7, 12, 13, 14, 15, 17, 18, 19, 21, 22, 29, 35, 37, 38, 39, and 40 contributed to the factor (lower than 0.300) with low correlation level. By removing these items and having known the factor structure of the scale in Gözütok et al. (2017) study, the analysis was repeated with two factors.

187

```
Variable  Real-data      Mean of random    95 percentile of random
          eigenvalues    eigenvalues       eigenvalues

  1       13.14182*      1.68374           1.78793
  2        1.56820       1.57285           1.64239
  3        1.22047       1.49474           1.55450
  4        1.09954       1.42710           1.48387
  5        0.86619       1.36827           1.41652
  6        0.76686       1.30996           1.36029
  7        0.73194       1.26011           1.30579
  8        0.66534       1.20939           1.25088
  9        0.60133       1.16456           1.20487
 10        0.57966       1.12175           1.15873
 11        0.56325       1.07735           1.11338
 12        0.50336       1.03517           1.07103
 13        0.44584       0.99469           1.02914
 14        0.42507       0.95507           0.98753
 15        0.36862       0.91563           0.95192
 16        0.35242       0.87684           0.91342
 17        0.33334       0.83872           0.87492
 18        0.29360       0.80076           0.83775
 19        0.26727       0.76125           0.79698
 20        0.24434       0.72384           0.76236
 21        0.23116       0.68334           0.72178
 22        0.19833       0.64217           0.68285
 23        0.16893       0.59847           0.63673
 24        0.15654       0.55234           0.59528
 25        0.11302       0.50139           0.54843
 26        0.09356       0.43049           0.48895

* Advised number of dimensions:  1
```
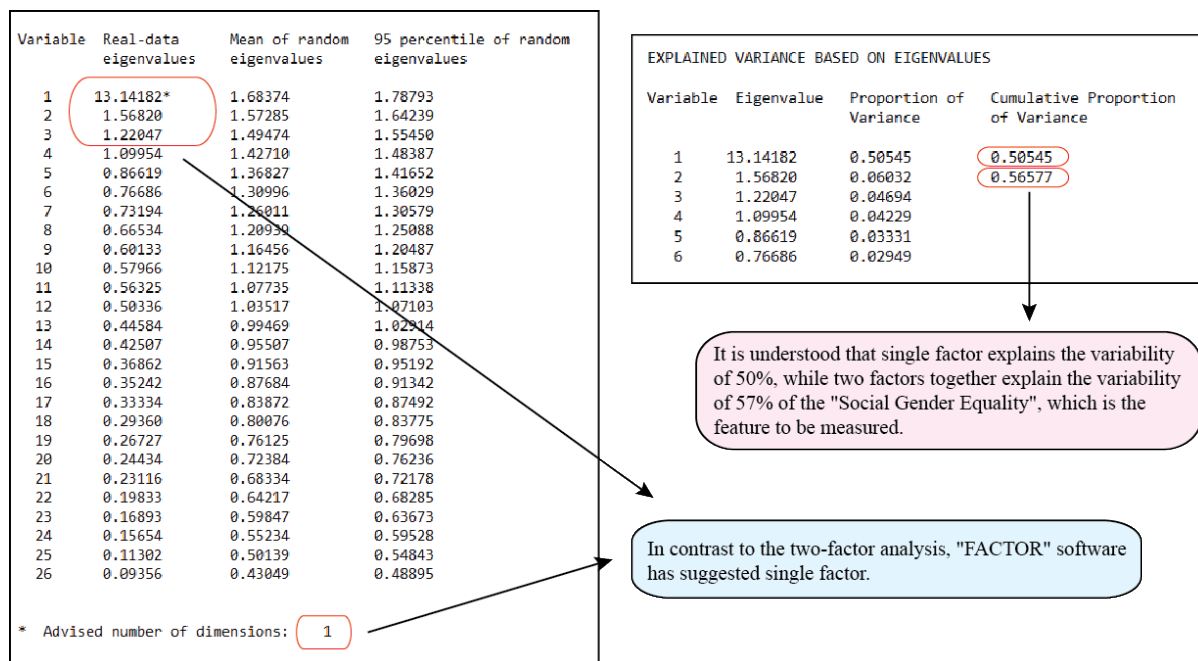
```
EXPLAINED VARIANCE BASED ON EIGENVALUES

Variable  Eigenvalue   Proportion of   Cumulative Proportion
                       Variance        of Variance

   1      13.14182     0.50545         0.50545
   2       1.56820     0.06032         0.56577
   3       1.22047     0.04694
   4       1.09954     0.04229
   5       0.86619     0.03331
   6       0.76686     0.02949
```

It is understood that single factor explains the variability of 50%, while two factors together explain the variability of 57% of the "Social Gender Equality", which is the feature to be measured.

In contrast to the two-factor analysis, "FACTOR" software has suggested single factor.

Figure 2. Analysis results obtained from "FACTOR" software by attempting two factors

As seen in Figure 2, the single factor, explained the 50% of the variability in SGE. The software suggested that analysis should be with single factor. The analysis was continued with a single factor structure. The results obtained were summarized in table 1.

Table 1.
Exploratory factor analysis and reliability analysis results based on polychoric correlation matrix

Item No	Item Loadings	Item Total Correlation	Item No	Item Loadings	Item Total Correlation
M4	-0.609	0.370	M27	0.796	0.633
M5	0.584	0.341	M28	0.847	0.718
M6	-0.692	0.478	M30	-0.711	0.505
M8	0.801	0.641	M31	-0.721	0.520
M9	-0.870	0.757	M32	0.704	0.496
M10	-0.765	0.586	M33	0.805	0.648
M11	0.594	0.353	M34	-0.581	0.337
M16	0.662	0.438	M36	-0.624	0.389
M20	-0.672	0.451	M41	0.769	0.592
M23	-0.622	0.387	M42	-0.549	0.302
M24	-0.572	0.327	M43	0.805	0.648
M25	-0.702	0.493	M44	0.827	0.684
M26	-0.696	0.485	M45	0.750	0.562

Kaiser Meyer Olkin (KMO) = 0.941
Bartlett = 4737,0, sd = 325, p<.01
Variance Explained by Single Factor = 0.505
Reliability = 0.961

When Table 1 is examined, it is seen that the loading values of the remaining 26 items do not fall below 0.549 and the item total correlation does not fall below 0.327. It has been seen that the items 4, 6, 9, 10, 20, 23, 24, 25, 26, 30, 31, 34, 36 and 42 have negative correlation values. In fact, some of these items are negative or have a negative meaning. Others do not have a negative meaning or are not negative. However, if the scale will become a single structure with all of the remaining materials after the analysis, it is possible that these items will be encoded negatively in order to make them meaningful.

KMO and Bartlett results of the structure in Table 1 show that it is within the accepted values in the literature. The value of reliability is at the desired level that is sought in social sciences. The variance

explained by the single factor is within the acceptable boundaries according to some authors (Büyüköztürk, 2013) while it cannot be accepted by others (Özdamar, 2013).

It is seen that Pearson's correlation matrix-based factor analysis results differ from the results based on the polychoric correlation matrix. Pearson's correlation matrix-based factor analysis results are summarized in Table 2.

Table 2.
Exploratory factor analysis and reliability analysis results based on Pearson correlation matrix

Item No	Factor	Item Total Correlation	Item Loadings After Rotation	
			Factor I	Factor II
M11	Factor 1	0,463	0,452	
M27	Factor 1	0,654	0,709	
M28	Factor 1	0,710	0,652	
M33	Factor 1	0,673	0,667	
M41	Factor 1	0,632	0,582	
M43	Factor 1	0,691	0,758	
M44	Factor 1	0,672	0,848	
M45	Factor 1	0,628	0,801	
M12	Factor 2	0,400		0,535
M14	Factor 2	0,479		0,619
M16	Factor 2	0,620		0,734
M22	Factor 2	0,333		0,632
M32	Factor 2	0,623		0,542

KMO = 0,922
Bartlett Sphericity (X2) = 2072,965; sd=78, p<0.01
Variance Explained by Factor 1 = %32,780
Variance Explained by Factor 2 = %20,051
Variance Explained by Both Factors Together= %52,831
Cronbach Alpha = 0,889

When the results in tables 1 and 2 are evaluated, it is seen that they are quite different. In the factor analysis based on the polychoric correlation matrix, items 4, 6, 9, 10, 20, 23, 24, 25, 26, 30, 31, 34, 36 and 42, which are shown with negative correlations, were removed from the scale in the Pearson correlation matrix-based factor analysis. Social gender equality's explained variance is close in both analyzes. KMO and Bartlett values and reliability values are at the desired level in both analyses (Büyüköztürk, 2013; Özdamar, 2013).

In accordance with the results obtained from the polychoric correlation matrix with the FACTOR software that are shown in table 1, the items 4, 6, 9, 10, 20, 23, 24, 25, 26, 30, 31, 34, 36, and 42 were reverse coded by returning to the same data set in the SPSS program. After this coding, in the SPSS software, additivity analysis was performed for 26 items and single factor structure that FACTOR software revealed. The results are summarized in table 3.

Table 3.
Results of additivity analyses of GES (Structure obtained by polychoric correlation matrix-based factor analysis)

Variance Source	Sum of Squares	Mean of Squares	F	sd	p
Nonadditivity	20.645	20.645	20.711	1	0.000

As a result of the analysis, the non-additivity state is meaningful in the structure consisting of 26 items and one dimension (Tukey Non-additivity p<.05). In this case, additivity is meaningless (Özdamar, 2013). In short, the structure is not in additivity state. The two-dimensional structure obtained by Gözütok et al. (2017) shows additivity. This difference can be explained by the fact that the Pearson correlation matrix accepts the data set at interval level and attempts to reach a scale with an additivity structure. However, the criterion of additivity for a scale structure of an ordinal measurement level is a weak one.

189

The structure obtained as a result of exploratory factor analysis based on the polychoric correlation matrix for GES is shown in table 1. For this structure, confirmatory factor analysis was applied with the unweighted least square (ULS) method. Unweighted least square (ULS) method is used for confirmatory factor analysis based on polychoric correlation (Katsikatsou, Moustaki, Yang-Wallentin, & Jöreskog, 2012). Confirmatory factor analysis with this method did not confirm the structure.

It has been determined that the two-dimensional structure obtained by Gözütok et al. (2017) has had close results with the fit indices obtained by Gözütok et al. (2017) when using confirmatory factor analysis by unweighted least square-ULS method. The results are summarized in table 4.

Table 4.
Fit indices after CFA

Estimation	χ^2	sd	χ^2/sd	RMSEA	AGFI	RMR	NFI
Pearson Maximum Likelihood	97,01	53	1,83	0,056	0,92	0,062	0,97
Unweighted Least Square	100,22	64	1,57	0,046	0,99	0,067	0,99

When table 4 is examined, it has been determined that the estimations made with the maximum likelihood and unweighted least square methods give very close fit index results. Koğar and Yılmaz Koğar (2015) stated that unweighted least squares (ULS) method used in confirmatory factor analysis of ordinal data is a suitable technique for estimating parameters with a minimum number of repetitions and estimating the parameters. Similar results have also been obtained in this study.

DISCUSSION and CONCLUSION

This study aims to compare the effect of using the Pearson or the polychoric correlation matrix on the analysis results in the process of developing the likert type scales in order to provide evidence of construct validity.

In studies conducted in the fields of social and behavioral sciences, all measurements have a degree of uncertainty/error regardless of precision and accuracy, which causes randomly and/or systematically, and consequently, these problems are reflected in the results. Especially in the analyses carried out while collecting evidence of reliability and validity of Likert type scales developed to be used as data collection instruments in scientific researches or adapted from one culture to another, it was found that using different correlation matrices could increase the error that would cause confusions in the results, and they may lead to a misinterpretation of the construct validity evidence, which is regarded as the most important of the validity types by many researchers (Cronbach & Meehl, 1955; Cronbach, 1984; Messick, 1988, 1993; Şencan, 2005; Urbina, 2014). Moreover, this situation will inevitably effect the results of subsequent studies carried out using the corresponding scales.

Similar to the previous studies carried out with simulated data (Holgado-Tello et al., 2010), it is seen in this real data used study that different correlation matrix based factor analyses results were found to differ from each other. In comparison to the previously held studies carried out with simulated data, this study is based on real data, and it also handles new discussions on number of items, items 'additivity states. As such, it differs from the previously held studies.

The findings of this study can be summarized as follows: As a result of the analyses carried out using the Polychoric or Pearson correlation matrix,

1) The items showed different "total item correlation" and "loading values".
2) The items could show a different (negative or positive) correlation coefficient.

190

3) Different factor numbers emerged.

4) Whether or not the resulting sub-dimensions' (factors) possibility of being additive or non-additive states have changed.

5) Different items were removed out of the scale.

6) Being confirmed and not confirmed status of the structure's being tested, has changed.

7) It was determined that estimations made by the maximum likelihood and unweighted least square methods have given very close fit index results.

All of these results clearly show that when the analyses are carried out with the data obtained from the Likert type scales, the results differ greatly from each other when the analyses are conducted by grounding Pearson correlation matrix at the interval measurement level or by grounding the polychoric correlation matrix, in which the data are accepted at the ordinal measurement level. From the perspective of the differences revealed by this study, which has conducted in terms of factor analysis, especially specific to construct validity; for researchers, who will use Likert-type scale or scales in their studies as a means of data collection by developing, and/or adapting, it is recommended to use both Pearson or polychoric correlation matrices for aforementioned analyses, and to compare the results by taking into account the theoretical background of the psychological feature they aim to measure. It is better to decide on which results they will choose, and the final construct they will prefer according to this comparison. Yet, the existence of two different structures will cause the view that the instruments developed and/or adapted for measurement do not actually measure the claimed psychological property, but measure a different property, or cause significant errors in the measurement results.

In the literature, there is a variety of information about the minimum number of participants required for the factor analysis applications in the measurement instrument development process. For example, Kline (2005) recommended at least 100; Hutcheson and Sofroniou (1999) stated at least 150 to 300 and Cattell (1978) claimed the desirable N to be at least 250 participants. Comrey and Lee (1992) described 100 participants as poor, 200 as moderate, 300 as good, 500 as very good and 1000 and more as excellent. Cattell (1978) emphasized that the number of participants in the factor analysis should be 3 to 6 times the number of the items and Gorsuch (1974) stated that it should be at least 5 times. Everitt stated that the number of participants should be at least 10 times the number of items (as cited in Arrindell and van der Ende, 1985, p. 166). The data set used in this study can be said to be obtained from a sufficient number of individuals according to the literature. The results showed that when worked with the proper sample size suggested by the literature, the solutions with Pearson correlation matrix give proper results.

Even though a real data set was used, one limitation of this study is the fact that the results of the factor analysis of the correlation matrices were discussed from a single data set. Performing methodological comparisons over more real data sets will help to make a healthier decision about which of the Pearson or polychoric correlation matrices is proper. Therefore, it is recommended to repeat this methodological comparison with different data sets within the scope of different studies. When performing factor analysis, discussing the results of the data sets which are below the sufficient sample level, can also support to make healthier decisions about the effectiveness of the methods as well. In this sense, it is recommended to repeat similar comparisons over different sized data sets.

This study is only limited to some of the results of factor analysis. Apart from them, other issues that are not taken into account both within the scope of factor analysis (such as cross loadings of items, different estimation methods, different types of rotations, whether different number of categories in the response format differ the results or not, etc.), and within the scope of analyses to obtain evidence of reliability, which is another important psychometric property for psychological measurement tools, and also within the scope of all other correlation-based analyses whether they are effected or not according to the used correlation matrix, can be studied with new studies.

REFERENCES

Albayrak, A.S. (2006). *Uygulamalı çok değişkenli istatistik*. Ankara: Asil Yayın

Amerikan Eğitim Araştırmaları Birliği, Amerikan Psikoloji Birliği, Eğitim Ölçümleri Uluslararası Konseyi (1997). *Eğitimde ve psikolojide ölçme standartları*. (S. Hovardaoğlu and N. Sezgin Trans.). (1st Edition). Ankara: Türk Psikologlar Derneği Yayınları. No 14

Anastasi, A. (1988). *Psychological testing* (6th Ed.). New York: McMillan.

Arrindell, W. A. & Van der Ende, J. (1985). An empirical test of the utility of the observations to variables ratio in factor and components analysis. *Applied Psychological Measurement*, 9(2), 165-178.

Atılgan, H., Kan, A., & Doğan, N. (2006). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık

Babakus, E., Ferguson, C. E., Jr., & Jöreskog, K. G. (1987), The Sensitivity of Confirmatory Maximum Likelihood Factor Analysis to Violations of Measurement Scale and Distributional Assumptions, *Journal of Marketing Research, 24* (May), 222- 228.

Bowden, S. C. (2004). The role of factor analysis in construct validity: Is it a myth? *Journal of International Neuropsychological Society. 10*, 1018-1019

Bryant, F. B., Yarnold, P. R., & Michelson, E. A. (1999). Statistical methodology: VIII. Using confirmatory factor analysis (CFA) in emergency medicine research. *Academic Emergency Medicine, 6*(1), 54-56

Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi. 32*, 470-483

Büyüköztürk, Ş. (2013). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Yayıncılık

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

Comrey, A. L. & Lee, H. B. (1992). *A first course in factor analysis*. New Jersey: Lawrence Erlbaum Associates Inc.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. USA: Cengage Learning

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

Cronbach, L. (1984). *Essentials of Psychological Testing*. Harper and Row, New York

Çokluk, Ö., Şekercioğlu, G. and Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları*. Ankara: Pegem Akademi

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9*, 327-346

Erkus, A. (2003). *Psikometri üzerine yazılar* (1st Edition). Ankara: Türk Psikologlar Derneği Yayınları

Erkuş, A. (2013). *Davranış bilimleri için bilimsel araştırma süreci*. Ankara: Seçkin Yayıncılık.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299

Gorsuch, R. L. (1974). *Factor analysis* (First Edition). Philadelphia: W. B. Saunders Company

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

Gözütok, F. D., Toraman, Ç., & Acar Erdol, T. (2017). Toplumsal cinsiyet eşitliği ölçeğinin (TCEÖ) geliştirilmesi. *İlköğretim Online, 16*(3), 1036-1048. doi: 10.17051/ilkonline.2017.330240

Guilley, W., & Uhlig, G. (1993). Factor analysis and ordinal data. *Education, 114*(2), 258-264

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

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero- García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity, 44*, 153-166

Hutcheson, G. & Sofroniou, N. (1999). *The multivariate social scientist, introductory statistics using generalized linear models*. The USA: Sage Publications

Jamieson, S. (2004). Likert Scales: How to (Ab)use them. *Medical Education, 38*, 1217-1218

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches, *Multivariate Behavioral Research, 36*(3), 347-387.

Jöreskog, K. G., & Sörbom, D. (2002). *Prelis 2: User's reference guide*. Lincolnwood: Scientific Software International, Inc

Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., & Joreskog, K. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics and Data Analysis, 56*(12), 4243-4258.

Kieffer, K. M. (1999). An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. *Research in the Schools, 6*, 75-92

Kline, T. J. B. (2005). *Psychological testing, a practical approach to design and evaluation*. The USA: Sage Publications

192

Koğar, H., & Yılmaz Koğar, E. (2015). Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Journal of Measurement and Evaluation in Education and Psychology, 6*(2), 351-364

Lemke, E. & Wiersma, W. (1976). *Principles of Psychological Measurement*. Chicago: Rand McNally.

Lorenzo-Seva, U., & Ferrando, P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*, 88-91

Messick, S. (1988). *The once and future issues of validity*: *Assessing the meaning and consequences of measurement*. In H. Wainer, & H. I. Braun. (Eds.), Test validity. (s. 33-48). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Messick, S. (1993). *Foundations of validity: Meaning and consequences in psychological assessment*. Presented at the Second Conference of the European Association of Psychological Assessment. Groningen, Netherlands

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research, 22*, 267-305

Nunnaly, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (Third Edition). NY: McGraw-Hill, Inc.

Öner, N. (1987). Kültürlerarası ölçek uyarlamasında bir yöntembilim modeli. *Türk Psikoloji Dergisi, 6*(21), 80-83

Özdamar, K. (2013). *Paket programlar ile istatistiksel veri analizi (Volume 1 and 2)* (9th Edition). Eskişehir: Nisan Kitabevi

Pedhazur, E. J., Pedhazer Schmelkin, L. (1991). *Measurement, design and analysis: An integrated approach* (1st Edition). NJ: Lawrence Erlbaum Associate, Inc

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 6*, 559-572

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201-293

Stapleton, C. D. (1997). *Basic concepts in exploratory factor analysis (EFA) as a tool to evaluate score validity: A right-brained approach* (pp. 1–8). Presented at the Southwest Educational Research Association, Austin

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680

Şencan, H. (2005). Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik. (1st Edition). Ankara: Seçkin Yayınları.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). USA: Pearson

Thomas, H. (1982). IQ interval scales, and normal distributions. *Psychological Bulletin, 91*, 198-202

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*, 197-208

Uebersax, J. S. (2015). *The tetrachoric and polychoric correlation coefficients. Statistical Methods for Rater Agreement*. Retrieved from http://john-uebersax.com/stat/tetra.htm

Urbina, S. (2014). *Essentials of psychological testing* (Second Edition). NJ: Wiley & Sons, Inc.

TÜRKÇE GENİŞLETİLMİŞ ÖZET

Psikolojik ölçme araçları, genelde insan davranışlarını konu edinmiş pek çok sosyal bilim dalında, özellikle psikoloji, eğitim, sosyoloji ve psikiyatride kullanılır. Genel anlamı ile psikolojik testler bireylerin yetenekleri, becerileri, performansları, güdüleri, tutumları hakkında bilgi veren ve bunlardan elde edilecek verilere göre çeşitli kararların alınmasına yardımcı olan sistematik yaklaşımlardır.

Türkiye'de değişik amaçlarla geliştirilmiş ya da uyarlanmış çok sayıda ölçme aracı bulunmaktadır. Yapılacak yeni araştırmalarda kullanılmak üzere, var olanlara ek olarak özellikle farklı psikolojik özellikleri ölçen, farklı örneklemler, farklı yaş grupları vb. için kullanılabilecek yeni ölçme araçlarına da ihtiyaç duyulmaktadır. Bu ihtiyaç ise, yeni ölçme araçlarının geliştirilmesi ya da uyarlanması yolu ile karşılanmaktadır. Her iki yöntemin de avantajlı ve dezavantajlı yönleri bulunmakla birlikte, asıl önemli olan bu araçlar için gerekli geçerlik, güvenirlik ve hatta mümkünse standardizasyon çalışmalarının uygun yöntemlerle ve bu yöntemlerin de doğru süreçler izlenerek yapılmış olmasıdır. Çünkü yapılan bir bilimsel çalışmanın teorik altyapısı ne kadar sağlam olursa olsun veri toplamak amacıyla kullanılan ölçme araçları gerekli psikometrik nitelikleri taşımıyorsa o çalışma ile ulaşılacak bulguların yorumlanmasında güven problemi olacak, bu araçların uygulanmasından elde edilen sonuçlarla yanlış kararlar alınması ise kaçınılmaz olacaktır. Ölçek geliştirme ya da uyarlama çalışmalarında, kuşkusuz en önemli adımlardan biri, geliştirilen ya da uyarlanan aracın psikometrik niteliklerine dair deneysel geçerlilik kanıtlarının ortaya konmasıdır. Bu bağlamda, söz konusu aracın ölçmeyi amaçladığı özelliği tam ve doğru bir şekilde ölçüp ölçmediğine ilişkin bir sorgulamanın yapılması gerekmektedir. Ölçeğe ait diğer psikometrik niteliklerin belirlenmesi bu sorgulamanın ardından yapılabilecek ve ölçeğin uygulanmasından elde edilecek puanlara dair sağlıklı kararlar da ancak, aracın geçerliliğine ilişkin kanıtlar doğrultusunda alınabilecektir.

Her ne kadar kapsam, yordama, zamandaş/eş-zamanlı ve yapı geçerliliği olmak üzere 4 kategori altında tanımlanmış olsa da günümüze kadar süre gelen tartışmalar içerisinde geçerlik kavramının, yapı geçerliği çatısı altında toplanması gerektiğini savunan araştırmacılar olmuştur (Cronbach ve Meehl, 1955; Cronbach, 1984; Messick, 1988, 1993; Şencan, 2005; Urbina, 2014). Birçok araştırmacı tarafından da yapı geçerliği, bu tür geçerliğe ait kanıt elde etmek amacıyla en sık kullanılan yöntem olan faktör analiziyle ilişkilendirilmiş, hatta yapı geçerliliği kavramı yerine faktör geçerliliği kavramını kullanmışlardır (Nunnally ve Bernstein, 1994).

Davranış bilimlerinde ölçek geliştirme ve uyarlama çalışmalarında; yapı geçerliğine ilişkin kanıtların ortaya konması, faktör yapılarının ortaya çıkarılması ya da daha önceden kestirilen faktör yapılarının doğrulanması amacıyla Nunnally ve Bernstein'nin (1994, sf. 111), psikolojik yapıların ölçümünün kalbi olarak tanımladığı, çok değişkenli istatistiksel yöntemlerden biri olan Faktör Analizi (FA) kullanılır. FA, diğer çok değişkenli istatistiklerle birlikte karmaşık araştırma sorularının ele alındığı çok sayıda bilimsel çalışmada kullanılmaktadır. Elbette eğitim, psikoloji, sosyoloji gibi sosyal bilim alanlarında da birden çok değişkenin eşzamanlı çalışıldığı ve tek değişkenli istatistiksel işlemlerin açıklamakta yetersiz kaldığı durumlarda; çok değişkenli varyans analizi, regresyon analizi, diskriminant analizi, çok boyutlu ölçekleme ve faktör analizi gibi çok değişkenli yöntemlere ihtiyaç duyulmaktadır. Bu yöntemlerin bir arada, bilimsel araştırmalarda ele alınan konuların kavramsal temellerinin ve nedensellik ilkelerinin anlaşılmasına açılan kemer şeklinde bir yapı oluşturduğunu varsayarsak, bu yapının kilit taşının "faktör analizi" olduğu söylenebilir. Tüm istatistiksel yöntemlerde olduğu üzere faktör analizi için de analiz öncesinde sorgulanması gereken hususların en başında verilerin hangi ölçek düzeyinde toplandığı gelmektedir. Sürekli değişkenlerle faktör analizi için kuramsal dayanaklar ve kullanılan yöntemler oldukça gelişmiştir. Ancak pratikte, gözlenen ve/veya ölçülen değişkenlerin bir kısmı sıralama ölçeği düzeyindedir ve bu durum sıklıkla göz ardı edilmekte ve doğru olmayan bir şekilde 1, 2, 3, 4 gibi sıralı kategorileri temsil eden sayılara metrik özelliğe sahipmiş gibi davranılmaktadır. Özellikle sosyal bilimler kapsamındaki tüm alanlarda oldukça sık kullanılmakta olan ölçeklerin (likert tipi vb.) madde

194

cevap formatlarından kaynaklı olarak, bu ölçeklerden elde edilen verilerin eşit aralıklı ölçek düzeyinde kabul edilerek buna uygun istatistiksel tekniklerin kullanılması, önceden beri eleştirile gelmiş bir durumdur. Bu eleştirileri ortaya koyan araştırmacılar söz konusu ölçeklerden elde edilen verilerin sıralama ölçeği düzeyinde olduklarını ve bu veriler için uygun istatistiksel tekniklerin kullanılması gerektiğini belirtmektedirler. Sıralama ölçeği düzeyinde değişkenler söz konusu olduğunda bu değişkenler arasındaki ilişkinin tahmin edilmesinde ya da bu değişkenler kullanılarak yapılacak ilişki temelli istatistiksel analizlerde "tetrakorik" (iki kategorili veriler için) ya da "polikorik" (üç veya daha çok kategorili veriler için) korelasyon matrisleri kullanılması gerekliliği vurgulanmaktadır. Faktör analizi kapsamında en tutarlı, en sağlıklı ve en güçlü kestirimlerin polikorik korelasyon matrisi kullanılarak yapılabildiği ortaya konulmuştur. Ancak ölçek geliştirme ve uyarlama süreçlerinde kullanılan hem keşfedici hem de doğrulayıcı faktör analizinin uygulanması sırasında genellikle Pearson korelasyon matrisinden faydalanılmaktadır. Ölçme aracının yapı geçerliğinin test edilmesi amacıyla faktör analizi kullanılacağı zaman, verilerin toplandığı ölçek düzeyinin göz önünde bulundurulması büyük önem taşımaktadır. Diğer taraftan analizlerde Pearson korelasyon matrisinin sıklıkla kullanılması, araştırmacıların konuyla ilgili yeterli bilgiye sahip olmamasından ya da kullanmakta oldukları bilgisayar paket programlarının sınırlılığından kaynaklanmaktadır.

Bu araştırmanın amacı, sosyal bilimlerde geliştirilen likert türü ölçeklere, ölçek geliştirme sürecinde yapı geçerliği kanıtı sağlamak amacıyla uygulanan faktör analizi kapsamında Pearson ya da polikorik korelasyon matrisi temelli analizin gerçekleştirilmesinin, analiz sonuçları üzerindeki etkisini karşılaştırmaktır.

Faktör analizinde Pearson ya da polikorik korelasyon matrisi kullanmanın analiz sonuçları üzerindeki etkisini incelemeyi ve analiz sonuçlarını karşılaştırmayı temel alan bu araştırma betimsel türde bir araştırmadır. Betimsel araştırmalar ilişkiyi ya da farkı merak etmeyen, neyin ne olduğunu saptamaya dönük çalışmalardır. Bu bakımdan betimsel çalışmalar, bilimin betimleme amacına hizmet etmekte ve aynı zamanda sonraki araştırmalar için denence üretmeye yönelik öngörü sağlarlar.

Sosyal ve davranış bilimleri alanlarında yapılan çalışmalarda ölçümlere sıklıkla rastgele ve/veya sistematik hatalar karışmakta, buna bağlı olarak ulaşılan sonuçlara da bu sorunlar yansımaktadır. Özellikle bilimsel araştırmalarda veri toplama aracı olarak kullanılmak üzere geliştirilen veya farklı bir kültürden bir diğerine uyarlanan Likert tipi ölçeklerin geçerlik ve güvenirlik kanıtları toplanırken yürütülen analizlerde, farklı korelasyon matrisi kullanımının sonuçlara karışacak hatayı artırabileceği ve bunun da en başta bir çok araştırmacı tarafından geçerlik türlerinin en önemlisi olarak görülen yapı geçerliği kanıtlarının yanlış yorumlanmasına yol açabilecektir. Dahası bu durum, söz konusu ölçekler kullanılarak yürütülecek sonraki çalışmaların sonuçlarına da kaçınılmaz olarak yansıyacaktır.

Daha önce yapılan ve simülatif verilerle yürütülen çalışmalarda olduğu gibi, gerçek verilerin kullanıldığı bu çalışmada da farklı korelasyon matrisi temelli faktör analizi sonuçlarının birbirinden farklılaştığı görülmüştür. Daha önce yapılan ve simülatif verilerle yürütülen çalışmalarda (Holgado-Tello ve diğerleri, 2010) olduğu gibi, gerçek verilerin kullanıldığı bu çalışmada da farklı korelasyon matrisi temelli faktör analizi sonuçlarının birbirinden farklılaştığı görülmüştür. Bu çalışma kapsamında ulaşılan bulgular şu şekilde özetlenebilir: Polikorik ya da Pearson korelasyon matrisi kullanılarak yürütülen analizler sonucunda,

1) Maddeler farklı "madde toplam korelasyonu" ve "yük değeri" göstermiştir.
2) Maddeler farklı yönde (negatif ya da pozitif) korelasyon değeri gösterebilmiştir.
3) Farklı faktör sayıları ortaya çıkmıştır.
4) Ortaya çıkan alt boyutların (faktör) toplanabilir olup olmama durumları değişmiştir.
5) Farklı maddeler ölçek dışı bırakılmıştır.
6) Test edilen yapının doğrulanıp doğrulanmama durumu değişmiştir.
7) Maximum likelihood ve unweighted least square yöntemi ile yapılan tahminlemelerin birbirine çok yakın uyum indeksi sonuçları verdiği belirlenmiştir.

195