International Journal of Assessment Tools in Education
2025, Vol. 12, No. 1, 113–130
https://doi.org/10.21449/ijate.1481201
journal homepage: https://dergipark.org.tr/en/pub/ijate **Research Article**

Factor extraction in exploratory factor analysis for ordinal indicators: Is principal component analysis the best option?

Tugay Kaçak[1*], Abdullah Faruk Kılıç[1]

[1]Trakya University, Faculty of Education, Department of Educational Sciences, Edirne, Türkiye

ARTICLE HISTORY
Received: May 09, 2024
Accepted: Nov. 04, 2024

Keywords:
PCA,
Factor extraction,
Monte Carlo simulation.

Abstract: Researchers continue to choose PCA in scale development and adaptation studies because it is the default setting and overestimates measurement quality. When PCA is utilized in investigations, the explained variance and factor loadings can be exaggerated. PCA, in contrast to the models given in the literature, should be investigated in categorical/ordered, severely skewed data, and multidimensional structures. The purpose of this study is to compare the relative bias and percent correct estimation of PCA, PAF, and MINRES techniques with Monte Carlo simulations. In Monte Carlo simulations sample size, level of skewness, number of categories, average factor loadings, number of factors, level of inter-factor correlation and test length were manipulated. The results show that PCA overestimates most models with lower average factor loadings, but PAF and MINRES provide unbiased results even with low factor loadings. PAF and MINRES produce more accurate and impartial results, and it is projected that PCA will lead researchers to believe that the items in scale development or adaptation studies are of "high quality."

1. INTRODUCTION

Factor analysis is frequently used as evidence of construct validity in scale development and adaptation studies. Several studies in the literature have examined how often researchers who develop or adapt scales use Principal Component Analysis [PCA] (Ford *et al.*, 1986; Gaskin & Happell, 2014; Goretzko *et al.*, 2019; Henson & Roberts, 2006; Koyuncu & Kılıç, 2019). The result of all these review studies is that PCA is frequently used in research. Despite the popularity of PCA, it is not recommended for the factor extraction step in EFA (Fabrigar *et al.*, 1999; MacCallum & Tucker, 1991). Although there are many methods for factor extraction, the attention given to PCA in scale development and adaptation studies makes investigating its usage particularly important. One of the main focuses of this study is to explore how PCA interacts with different data characteristics and to compare it with other widely recognized and robust methods. In addition to determining the performance of methods, it is also necessary to demonstrate their implications for empirical studies. In this study, as we have anticipated, we hope that studies examining the performance of methods can provide valuable insights for method

*CONTACT: Tugay KAÇAK tugaykacak@trakya.edu.tr Trakya University, Faculty of Education, Department of Educational Sciences, Edirne, Türkiye

e-ISSN: 2148-7456

selection in empirical research. We reviewed studies indexed in two different databases and analyzed the stages of scale development/adaptation studies, compiling the characteristics of the data. With this line, we aimed to highlight the critical steps of EFA and data characteristics, focusing on factor extraction methods in empirical studies. The study includes a systematic review followed by a series of Monte Carlo simulations designed to critique the findings derived from this review.

We first examined the use of exploratory factor analysis (EFA) in journals indexed in TRIndex (Türkiye) and the Web of Science (journals in the Q1, Q2, and Q3 quartiles of the SSCI) in terms of the methods used as factor extraction in EFAs. We have compared the studies indexed in WoS and TRIndex to provide Turkish researchers with a perspective. Then, we compared the estimation performance of PCA with minimum residual (MINRES) and principal axis factoring (PAF) in a simulation study. To determine the use of factor extraction methods in studies published in Türkiye and internationally, we examined the studies searched in TRIndex and WoS (Q1, Q2, and Q3) between 2015 and 2023. Koyuncu & Kılıç (2019) focused on the studies about social sciences were published between 2006-2016. We aimed to reveal the current usage of EFA, specifically factor extraction methods. Thus, the inclusion criterion was specified as being published between 2015 and 2023. In addition, we focused on scale development and scale adaptation studies published in the field of education.

We searched with the keywords "scale development, exploratory factor analysis, factor analysis, validity.". We used “published in journals indexed in TRIndex”, “Published in the field of “education.”, and “Published in journals (Q1, Q2 or Q3) indexed in WoS” as inclusion criteria, “Studies related to nursing, engineering, and training sciences were not included in the study to show similarities with the fields of the studies in the TRIndex.”, and “Studies in journals indexed in both TRIndex and WoS are considered in the WoS category and are not included in the TRIndex category.” as exclusion criteria.

As a result of the searches with keywords, 675 studies in journals indexed in TRIndex and 819 studies in journals indexed in WoS (Q1, Q2, and Q3) were found. For each search group, 100 studies were randomly selected and reviewed. The findings of the studies reviewed within the scope of the research are presented in Table 1.

Table 1. *Review of articles which use EFA.*

Number of Categories	TRIndex (n = 100)	WoS (n=100)	Factor Extraction Methods	TRIndex (n = 100)	WoS (n=100)
2	0%	1%	PCA	67%	39%
3	4%	4%	PAF	4%	29%
4	3%	8%	ML	3%	15%
5	89%	57%	ULS	0%	3%
6	0%	7%	MINRES	0%	2%
7	1%	18%	IMAGE	0%	2%
Not specified	2%	0%	WLSMV	1%	1%
Others	1%	5%	Not specified	25%	8%
			FIML	0%	1%
Sample Size	**TRIndex**	**WoS**	**Factor Rotation Methods**	**TRIndex**	**WoS**
0-99	1%	3%	Varimax	59%	30%
100-199	8%	20%	Promax	4%	22%
200-299	20%	23%	Direct Oblimin	6%	25%
300-399	33%	17%	Oblique(?)	1%	11%
400-499	15%	13%	Geomin	1%	3%
≥500	23%	24%	Promin	1%	1%
Mean	418	569	Equamax	1%	0%
			Not specified/ Unrotated	27%	8%

Number of Factors (p)	TRIndex	WoS	Mean of Factor Loadings	TRIndex	WoS
1	14%	8%	$0.4 \leq \lambda < 0.6$	22%	10%
2-3	30%	31%	$0.6 \leq \lambda < 0.8$	76%	85%
$p \geq 4$	56%	61%	$\lambda \geq 0.8$	0%	3%
Item(k)/Factor(p) Ratio	TRIndex	WoS	Others	2%	2%
$k/p \leq 3{:}1$	0%	0%	Mean	0.652	0.683
$3{:}1 < k/p \leq 5{:}1$	16%	46%			
$5{:}1 < k/p \leq 10{:}1$	67%	46%			
$k/p > 10{:}1$	17%	8%			
Number of Variables	TRIndex	WoS	Inter-factor Correlations	TRIndex	WoS
$k \leq 10$	3%	8%	$\varphi > \lvert 0.30 \rvert$	3%	26%
11-20	26%	35%	Including $\varphi < \lvert 0.30 \rvert$ correlations	4%	21%
21-30	36%	32%	Not reported (Oblique)	10%	22%
$k \geq 31$	35%	25%	Uncorrelated factors or unidimensional structure	83%	31%

1.1. Number of Categories

For the studies published in journals indexed in both TRIndex and WoS, it is seen that the majority of them were developed in 5-point Likert type (73%). 2 studies indexed in TRIndex determined that only Likert-type scales were used. However, no information was provided about the number of categories. Since it will also change the type of correlation matrix to be created according to the number of categories of the data, it affects the analysis processes (Holgado–Tello *et al.*, 2010).

1.2. Factor Extraction Methods: PCA vs the Others

PCA was the most frequently used factor extraction method in the studies searched in TRIndex and WoS (53%). It was determined that 25 studies in TRIndex and 8 in WoS (17%) did not report factor extraction methods. Factor extraction methods must be reported due to their assumptions and performance under various conditions (Goretzko *et al.*, 2019). There are studies in literature that compare factor extraction methods under various conditions. Although it is the most frequently used factor extraction method, studies indicate that PCA is not a factor analysis method (Fabrigar *et al.*, 1999; Harman, 1970; MacCallum & Tucker, 1991; Mulaik, 1990). In addition, Matsunaga (2010) states in his study that PCA can not be used instead of exploratory factor analysis methods because it determines the components by taking the diagonal elements in the correlation matrix with a value of 1.00 - that is, perfect reliability - without including the error variance.

In contrast to these views, studies argue that PCA is preferable (Arrindell & Van Der Ende, 1985; Costello & Osborne, 2005). Although there is no consensus on factor extraction methods, the current literature recommends using factor extraction methods that separate the error variance. Therefore, examining the performance of factor extraction methods will enlighten practical applications.

1.3. Sample Size

Most sample sizes of randomly selected and reviewed studies are between 300-399 sample size range. For the studies searched in TRIndex, the average sample size is 418, the minimum sample size is 46, the maximum sample size is 2083, and the median is 351. For the studies indexed in WoS, the average sample size is 569, the minimum sample size is 55, the maximum sample size is 9231, and the median is 314.5. It is seen that the sample size of 84% (n=168) of the reviewed studies is larger than 200, which is the minimum sample size required for EFA, as stated in Fabrigar *et al.* (1999).

1.4. Inter-factor Correlations and Factor Rotation Methods: Oblique or Orthogonal Rotation?

It is seen that most of the studies consisted of at least four dimensions (56% for TRIndex, 61% for WoS). With these findings, Varimax rotation (makes the factors as uncorrelated) is the most popular rotation method. Although a large of number of multidimensional constructs, still orthogonal rotations were preferred. This two findings conflict with the literature about the construct of interest in social sciences which commonly are correlated and multidimensional.

Total score analyses should not be performed with multidimensional scales (ntotal=25) that are multidimensional and have correlations less than |.30|. Although there is no certainty that the correlation between factors will be significant and above .30 when oblique rotation is preferred, it is theoretically possible that the factors may be unrelated after oblique rotation. In this case, it will be necessary to examine the scale structure regarding reproducibility for the studies in which oblique rotation was preferred and did not report the correlation between factors (ntotal = 33). In addition, it was found that the Varimax rotation method was preferred for one-factor structures in 2 studies in TRIndex and 1 study in WoS, and rotation methods for single-factor structures are not theoretically appropriate (Osborne, 2015). Direct Oblimin (16%) was frequently used in the studies. Thirty-four studies (17%) did not report the rotation method. Since oblique rotation methods allow for all levels of correlation between factors, it is suggested to be used for related and uncorrelated constructs (Comrey & Lee, 1992; Costello & Osborne, 2005; Nunnally & Bernstein, 1994).

1.5. Number of Variables, Factors and Items per Factor Ratio

None of the reviewed studies had a lower than 3:1 item/factor ratio recommended by Brown (2006) and Downing & Haladyna (2006) as the minimum ratio. In terms of the ideal ratio of items per factor, 5:1 (Gorsuch, 2015), 10:1 (Nunnally & Bernstein, 1994) has been suggested for EFA. In contrast to all of these, MacCallum *et al.* (2001) reject just one ratio criterion; they suggest focusing on the quality of items (factor loadings). The studies in TRIndex are mostly clustered in the 5:1 and 10:1 range, and in WoS, they are located mostly in the range of 3:1 between 5:1 and 5:1 between 10:1.

1.6. Factor Loadings

Numerous cut points about the factor loadings of variables can be found in the literature. In Hinkin's (1995) study 0.40, Costello and Osborne (2005) 0.30, Tabachnick and Fidell (2013) suggested that a loading of 0.32 would be significant. These cut points are towards the loadings of the variables on the primary factors. We evaluated the studies for average factor loadings as 0.40 low, 0.60 medium, and 0.80 high (Comrey & Lee, 1992). The average factor loadings are above 0.60 for studies in both groups. In the "Others" group, there are studies that did not publish factor loadings, reported factor loadings above a factor loading value, or published average factor loadings. Factor loadings should be reported as they provide information about the items' quality and the measurements' quality.

1.7. Current Study

PCA extracts the principal factors/components from the correlation matrix with diagonal elements of 1.00, and each extracted factor aims to explain the maximum amount of the correlation matrix that can be obtained. Since the diagonal elements do not change, this method tries to determine the entire variance for a variable. Unlike PCA, MINRES (equivalent to ULS according to Jöreskog (2003)) aims to maximally reproduce the off-diagonal elements in the correlation matrix using a least squares approach. This causes the operations performed on the correlation matrix to differ according to the methods. Therefore, the results obtained vary according to the methods. Although PCA practically takes the diagonal elements in the correlation matrix as 1, PAF differs from PCA in focusing on common variance (Mabel & Olayemi, 2020). Methods such as maximum likelihood (ML), alpha factoring, image factoring, and GLS, which follow different assumptions and procedures for factor extraction, are also available in the literature. Specifically, ML assumes multivariate normality (Garson, 2023) which is often violated by ordinal/categorical datasets (Kaplan, 2004). Fabrigar & Wegener (2012) discussed ML with ordinal/categorical datasets. Thus, it is clear that performance of ML is limitless with non-continuous datasets, and we decided to exclude ML. Watkins (2018) recommends PAF to deal with non-normal datasets. With this recommendation, PAF was the other method that we chose to analyze. Third method, MINRES, have no distributional assumptions (Jöreskog, 2003), so we decided to examine performance MINRES in this study. In sum, this study focuses on PAF, PCA, and MINRES for listed reasons. Previous studies have examined the PCA method, but its application to simulation studies typically involves continuous data sets. Therefore, in the current study, we performed analyses for the 5-point Likert type data set, a commonly used data set. Additionally, we examined binary data. Unlike other studies in literature, this study examined how biased the average factor loadings were. Therefore, it was possible to observe the practical outcomes of using PCA. Table 1 demonstrates the frequent use of PCA, despite its examination in previous studies. Therefore, this study stands out from others in literature and holds significant importance. Detailed information about other factor extraction methods will not be given. In addition, it can be said that PCA is frequently used among factor extraction methods because it overestimates factor loadings, explains total variance, and is set as default in most statistical software. Simulative studies examining the performance of the focused methods are given in Table 2.

The studies in Table 2 show that factor extraction methods have been examined under many conditions. These studies were mainly conducted with normally distributed continuous data sets. However, as accepted in educational research, the assumption that psychological characteristics are normally distributed is often not met due to the characteristics of the samples (Ho & Yu, 2015). In addition, indicators are mostly ordinal. Considering all these reasons, more work needs to be done for ordinal data with skewed distributions. Unlike earlier studies, this study focused on ordinal variables followed normal and non-normal distribution that are mostly encountered in educational and psychological structures.

Table 2. *Simulation studies in the literature.*

Studies	Factor Extraction Methods	Data Type	Distribution	Sample Size	Test Length	Number of Factors	Factor Loading / Communalities	Inter-factor correlation and rotation method
Widaman (1993)	PCA, PAF	Continuous	Normal	200	9, 18, 36, 24, 48, 96	3	0.40 0.60 0.80	None Varimax 0.50 Harris-Kaiser Orthoblique
Snook & Gorsuch (1989).	PCA, PAF	Continuous	Normal	200	9, 18, 36	3	0.40 0.60 0.80	None Varimax
De Winter & Dodou (2016).	PCA, PAF, ML	Continuous	Normal	50 5000	10, 50, 100	2,3,4,5	0.30 0.60	Varimax, Direct Quartimin, Procrustes Rotation
Coughlin (2013)	PAF, OLS, ML	Mixed (5%, 25%, 50%, 75%, 95%) Dichotomous and continuous	Normal	100, 200, 300,1000	20, 40, 60	2, 4, 8	High – 0.6, 0.7, 0.8 Wide – 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 Low - 0.2, 0.3, 0.4	Varimax
This study	MINRES, PCA, PAF	Ordinal (2 and 5)	Right-skewed, normal, left-skewed	200, 500, 1000	5 – 10 items per factor	1, 2, 3	0.40 0.70	0.00 0.30 0.60 Varimax and Promax

1.8. Importance

In this study, the performance of PCA in predicting factor loadings and inter-factor correlations is compared with MINRES and PAF. It can be said that this comparison will contribute to the literature in the following four aspects: 1) examining how accurate PCA, which is frequently used in scale development studies, gives accurate results will shed light on practitioners in practice; 2) the performance of PCA on categorical data can be examined in areas where categorical data are frequently used, 3) the performance of MINRES and PAF, which are recommended to be used on skewed data, can be examined on skewed data and their performance can be compared with PCA, 4) unlike other studies in the literature, the effects of categorical EFA on factor extraction can also be examined since it is studied with categorical data. Therefore, this study is important in providing information about the dominant use of PCA in the literature and the results obtained from the scales developed with PCA. In this direction, the study aims to investigate:

1. What is the average factor loading bias?
2. What is the percentage of correct estimation of average factor loading?

2. METHOD

A Monte Carlo simulation study examined which factor extraction method gives more accurate results for the examined models. The focus of the study was principal component analysis. Monte Carlo simulations are studies where data is produced according to a certain distribution, the produced data is analyzed with different statistical methods, and the results are compared (Sigal & Chalmers, 2016). We examined principal component analysis, principal axis, and MINRES methods.

2.1. Simulation Conditions and Data Generation

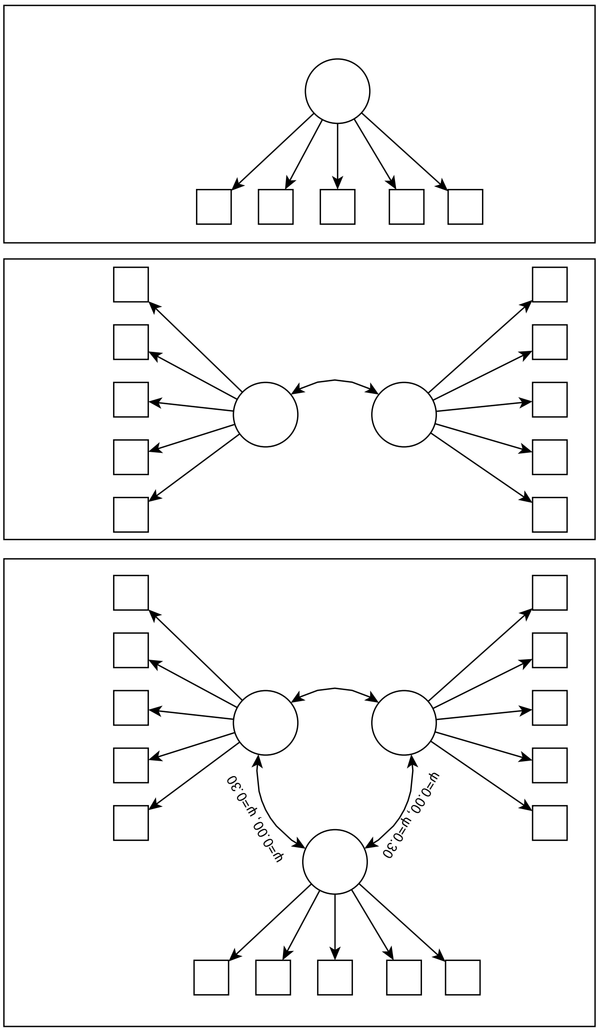
The current study examined the factor extraction method's performance; we determined the simulation factors as the number of categories, measurement model, items per factor, average factor loadings, distribution of variables, and sample size. Table 3 presents the simulation conditions.

Table 3. *Simulation conditions.*

Simulation Factors	Simulation Conditions	The number of conditions
Measurement Models (Figure 1)	Unidimensional, 2 factors (ψ=0.00), 2 factors (ψ=0.30), 3 factors (ψ=0.00) and 3 factors (ψ=0.30)	5
The number of categories	2 and 5	2
Items per factor	5 and 10	2
Average factor loadings	0.40 and 0.70	2
Distribution of variables	Left skewed, normal, and right skewed	3
Sample size	200, 500, and 1000	3
	Total	2x5x2x2x3x3=360 with 1000 replication

In the review study conducted by Koyuncu & Kılıç (2019), it was reported that more than half (55.5%) of the scale development studies in the field of social sciences had 1-3 dimensions. Therefore, the number of dimensions in the current study was determined to be 1, 2, and 3. Inter-factor correlations were determined to examine factor extraction methods' performance in unrelated and moderately related constructs. The number of categories of variables was manipulated as 2 and 5. Variables with two categories can be in achievement tests or checklists. Otherwise, measurement results in different fields, such as health, can also be categorical. For this reason, a 2-category condition was added to the study. On the other hand, since the most frequently used category number in Likert-type scales is 5 (Lozano *et al.*, 2008), it was added to the study.

Figure 1. *Measurement models.*



We manipulated items per factor as 5 and 10 items. Since it is known that there should be at least three items in a factor for a factor to be formed (Brown, 2006), the minimum number of items was considered to be 5 in this study. Considering the 3-dimensional structures, when items per factor are 10, the upper limit of the number of items is set as 10 since 30-item measurement tools will be formed. The average factor loading was manipulated to be 0.40 and 0.70. Although there are different suggestions for the minimum factor loading to be obtained as a result of EFA, it can be said that it will generally be around 0.30 (Costello & Osborne, 2005; Howard, 2016; Tabachnick & Fidell, 2013). Therefore, in this study, the average factor loading condition was determined to be 0.40. At the same time, it can be said that the factor loadings of the items will be higher in stronger scales. For this reason, the condition of 0.70 was added to the study to include scales with better items.

The skewness of the variables is also an issue that needs to be studied. Costello & Osborne (2005) states that the PAF method can be used when the variables are skewed. On the other hand, Zygmont and Smith (2014) stated that the MINRES method can be used in skewed variables. Therefore, in order to evaluate the performance of these methods on skewed data, both left and right-skewed conditions were added to the study. Normal conditions were also added to the study to evaluate the changes that may occur in the performance of the methods under conditions with normal distribution. The skewness coefficient of the variables was set as -2.5 for left-skewed variables and 2.5 for right-skewed variables. According to the literature of skewness, ±2 skewness can be justified as an acceptable limit for normality (Hair, 2014). Thus, ±2.5 skewness may be justified as non-normal, or skewed distributions.

At last, the sample size was manipulated to be 200, 500, and 1000. In studies in the literature, these sample sizes are often considered small, medium, and large. They are also frequently used in simulation studies (Beauducel & Herzberg, 2006; Kılıç & Doğan, 2021; Li, 2016; Oranje, 2003; West *et al.*, 1995) For this reason, sample sizes were handled in this way in this study.

2.2. Evaluation Criteria

We used relative bias (RB) and percent correct (PC) as evaluation criteria in the study. Relative bias is calculated as follows;

$$Relative\ Bias = \frac{\bar{\psi} - \psi_{True}}{\psi_{True}} \qquad 1$$

Where $\bar{\psi}$ is the average of the estimated factor loadings across 1000 replications, and ψ_{True} is the true average factor loading. |RB| > 0.10 means substantial bias (Flora & Curran, 2004; Forero *et al.*, 2009; Rhemtulla *et al.*, 2012).

We calculate the ±5% of the true factor loadings for percent correct. Then, for 1000 replications, we examined what proportion of the average factor loadings estimated by the models fell between this range (±5%). We used 90% PC value as "acceptable" in this study. (Collins *et al.*, 2001).

2.3. Data Analysis

We used a uniform distribution to determine factor loadings. First, we determined the factor loadings for the population, yielding an average factor loading of 0.40 and 0.70. Second, we generated continuous data followed by a multivariate normal distribution. Lastly, we categorized the dataset using predetermined thresholds. We used thresholds in Appendix 1.

We used the "lavaan" package (Rosseel, 2012) to generate data. The generated data sets were analyzed with the "psych" package (Revelle, 2024). Polychoric correlation matrices were used in the analysis. In multidimensional structures, Promax was used in conditions with an inter-factor correlation of 0.30, and Varimax was used in conditions with an inter-factor correlation of 0.00.

3. RESULTS

3.1. Relative Bias of Factor Loadings

One-way ANOVA was conducted to determine the simulation conditions influencing the RB values. ANOVA results indicated that all of the simulation conditions have an effect on the RB values. Partial eta squares of each simulation condition are represented in Table 4.

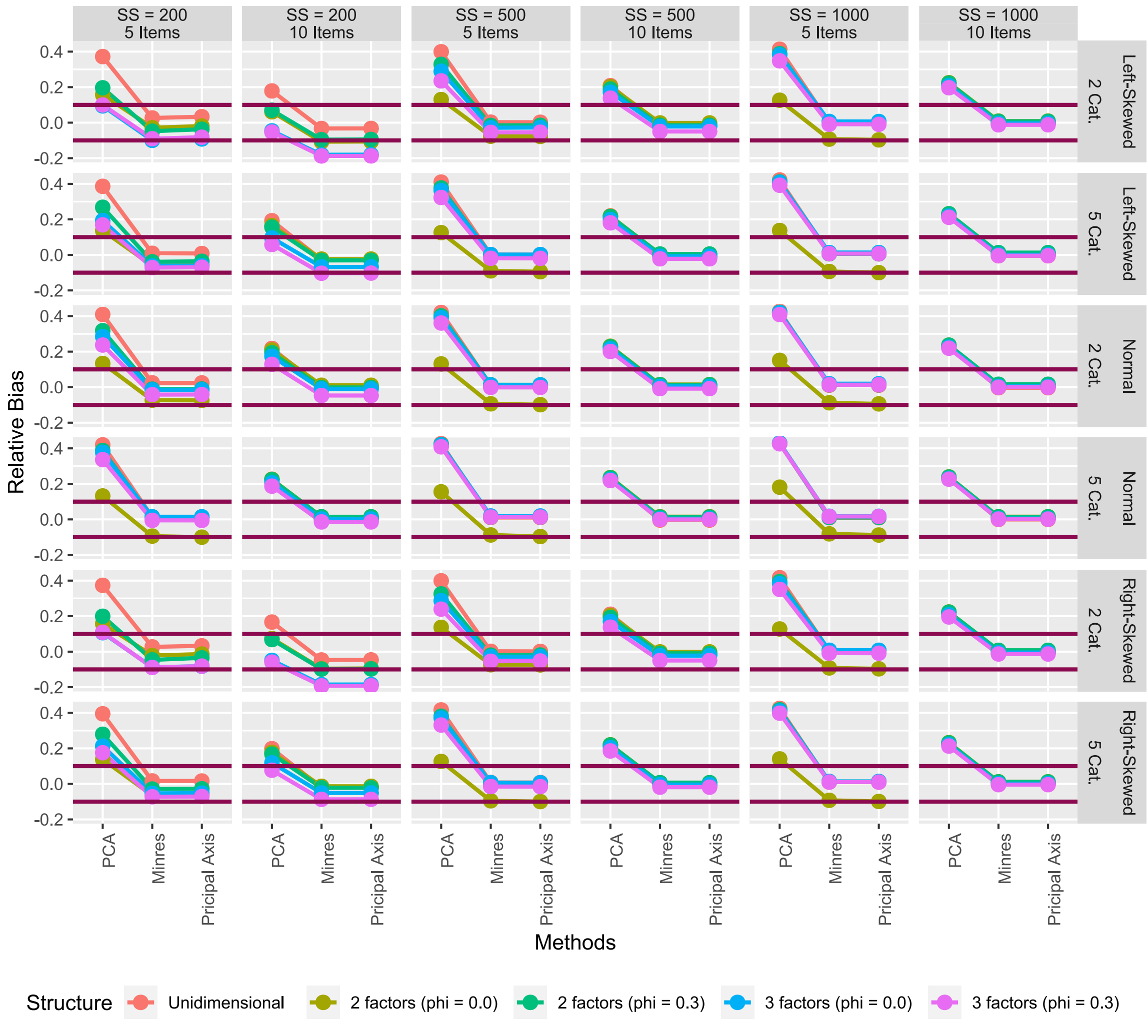
Table 4. *The ANOVA results for each simulation factors on RB values.*

Simulation Factors	The ANOVA Results
Measurement models	[$F(4, 1066) = 11.61, p < 0.01, \eta^2 = 0.04$]
The number of categories	[$F(1, 1066) = 11.86, p < 0.01, \eta^2 = 0.01$]
Items per factor	[$F(1, 1066) = 41.38, p < 0.01, \eta^2 = 0.04$]
Average factor loadings	[$F(1, 1066) = 122.86, p < 0.01, \eta^2 = 0.10$]
Distribution of variables	[$F(2, 1066) = 11.99, p < 0.01, \eta^2 = 0.02$]
Sample size	[$F(2, 1066) = 30.19, p < 0.01, \eta^2 = 0.05$]
Factor extraction method	[$F(2, 1066) = 803.29, p < 0.01, \eta^2 = 0.60$]

All simulation factors and factor extraction method have statistically significant effect on RB values. Measurement models [$F(4, 1066) = 11.61, p < 0.01, \eta^2 = 0.04$], the number of categories of variables [$F(1, 1066) = 11.86, p < 0.01, \eta^2 = 0.01$], items per factor [$F(1, 1066) = 41.38, p < 0.01, \eta^2 = 0.04$], average factor loadings [$F(1, 1066) = 122.86, p < 0.01, \eta^2 = 0.10$], distribution of variables [$F(2, 1066) = 11.99, p < 0.01, \eta^2 = 0.02$], sample size [$F(2, 1066) = 30.19, p < 0.01, \eta^2 = 0.05$], and factor extraction method [$F(2, 1066) = 803.29, p < 0.01, \eta^2 = 0.60$]. The mean and median of RB of the methods for all conditions were 0.16 and 0.10 for PCA, -0.01 and 0.00 for MINRES, and -0.01 and 0.00 for PAF, respectively.

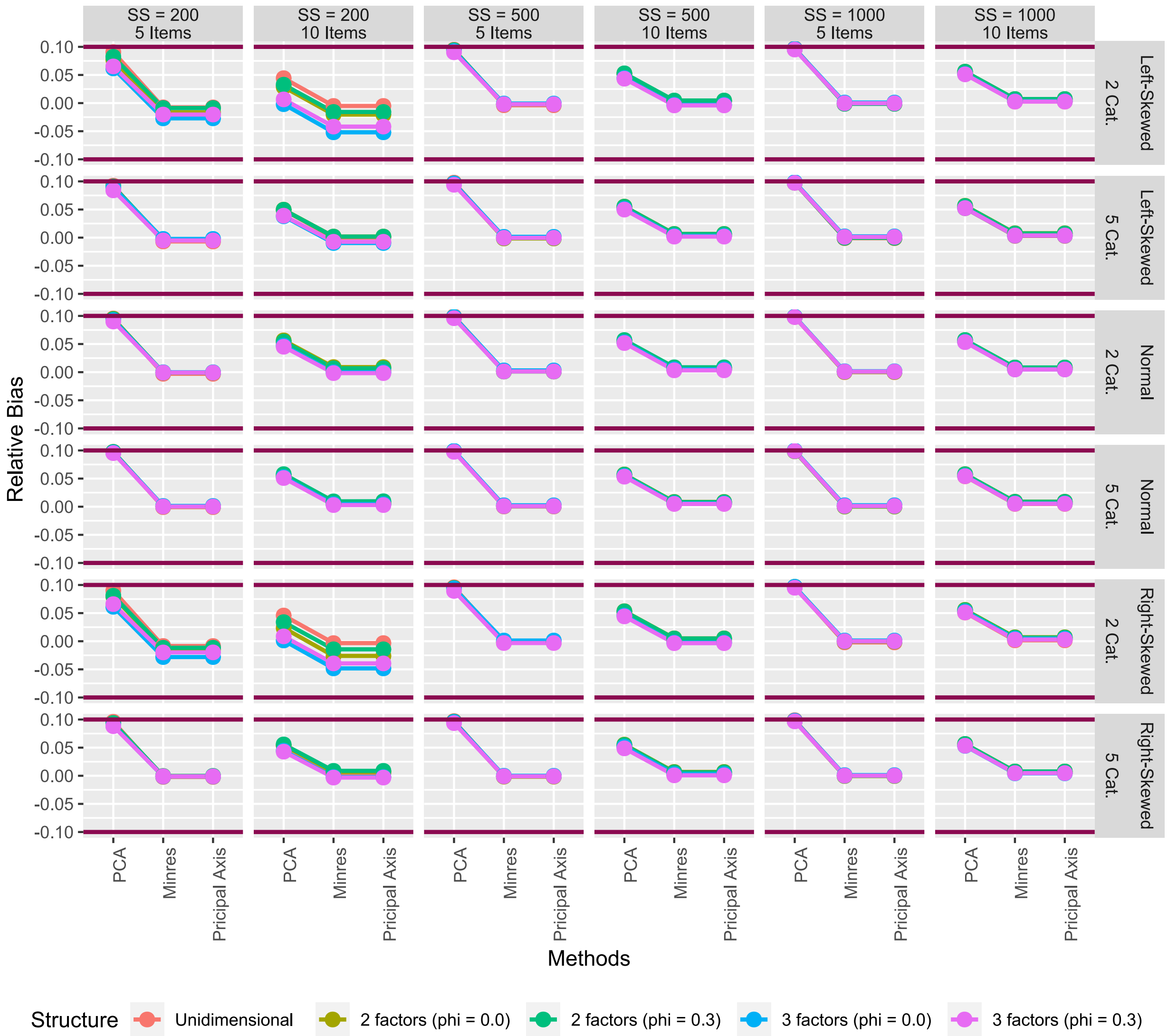
The graphs of the RB values are presented in Figures 2 and 3 with average factor loadings of 0.40 and 0.70, respectively. The relative bias values are within the acceptable range for PCA, MINRES and PA in all conditions where the average factor loading is 0.70 (see Figure 3).

Figure 2. *RB values of methods for average factor loading is 0.40.*



PCA is biased within the acceptable range in 10 (5.6%) conditions with an average factor loading of 0.40. When these ten conditions are analyzed, it can be said that the distribution is skewed, the sample is small, the item per factor is high, the structure is 3-dimensional, and the number of categories is low. In these conditions, while MINRES and PAF are under factoring, PCA is biased within the acceptable range. In other words, it can be said that there is a structure suitable for the general pattern. PCA overestimated the factor loading in all other conditions except for these conditions. MINRES underestimated in 7 (3.9%) conditions where the average factor loading was 0.40. These conditions were observed in cases where the number of items was high, variables were skewed, multifactor structures, and dichotomous variables. PAF underestimated factor loadings in 9 (5%) conditions where the average factor loading was 0.40.

Figure 3. *RB values of methods for average factor loading is 0.70.*



3.2. Percent Correct of Factor Loadings

One-way ANOVA was conducted to determine the simulation conditions influencing the PC values. ANOVA results indicated that all of the simulation conditions have an effect on the PC values. The mean of PC values statistically significantly differed from each other in terms of the number of categories of variables [$F(1, 1066) = 6.23$, p <0.05, η^2 =0.01], items per factor [$F(1, 1066) = 197.410$, $p < 0.01$, $\eta^2 = 0.16$], average factor loading [$F(1, 1066) = 967.06$, $p < 0.01$, $\eta^2 = 0.48$], structure [$F(4, 1066) = 8.27$, $p < 0.01$, $\eta^2 = 0.03$], distribution of variables [$F(2, 1066) = 20.98$, $p < 0.01$, η^2 =0.04], sample size [$F(2, 1066) = 96.79$, $p < 0.01$, $\eta^2 = 0.15$], and method [$F(2, 1066) = 1597.47$, $p < 0.01$, $\eta^2 = 0.75$]. The mean and median of the methods in terms of the PC values for all conditions were 14.70% and 2.9% for PCA, 72.82% and 80.45% for MINRES, and 72.79% and 80.45% for PAF, respectively. The graphs are presented in Figures 4 and 5 with average factor loadings of 0.40 and 0.70, respectively.

PC values are lower than 90% for most conditions of average factor loading, which is 0.40. In 15 conditions (8.33%), MINRES and PAF have PC values above 90%, while in the other conditions where the average factor loading is 0.40, they have PC values below 90%. When the 15 conditions with adequate performance are examined, it can be said that the sample is mostly 1000, the variable distribution is normal, the number of items is 10, the number of categories is five, and the number of dimensions is 3. In the same conditions, PCA could not reach 90%.

PC values increased for MINRES and PAF in conditions where the average factor loading was 0.70. PCA has PC values above 90% only in two conditions. These conditions were observed in data sets with a sample size of 200, 2 categories, 10 items, three factors, and skewed distribution. In conditions where the average factor loading was 0.70, MINRES and PAF failed to perform adequately in 51 conditions (28.33%). When these conditions were examined, it was observed that the sample was small, the variables were skewed, and they were in multidimensional structures. The number of items and categories does not affect the performance of the methods.

Figure 4. *PC values of methods for average factor loading is 0.40.*

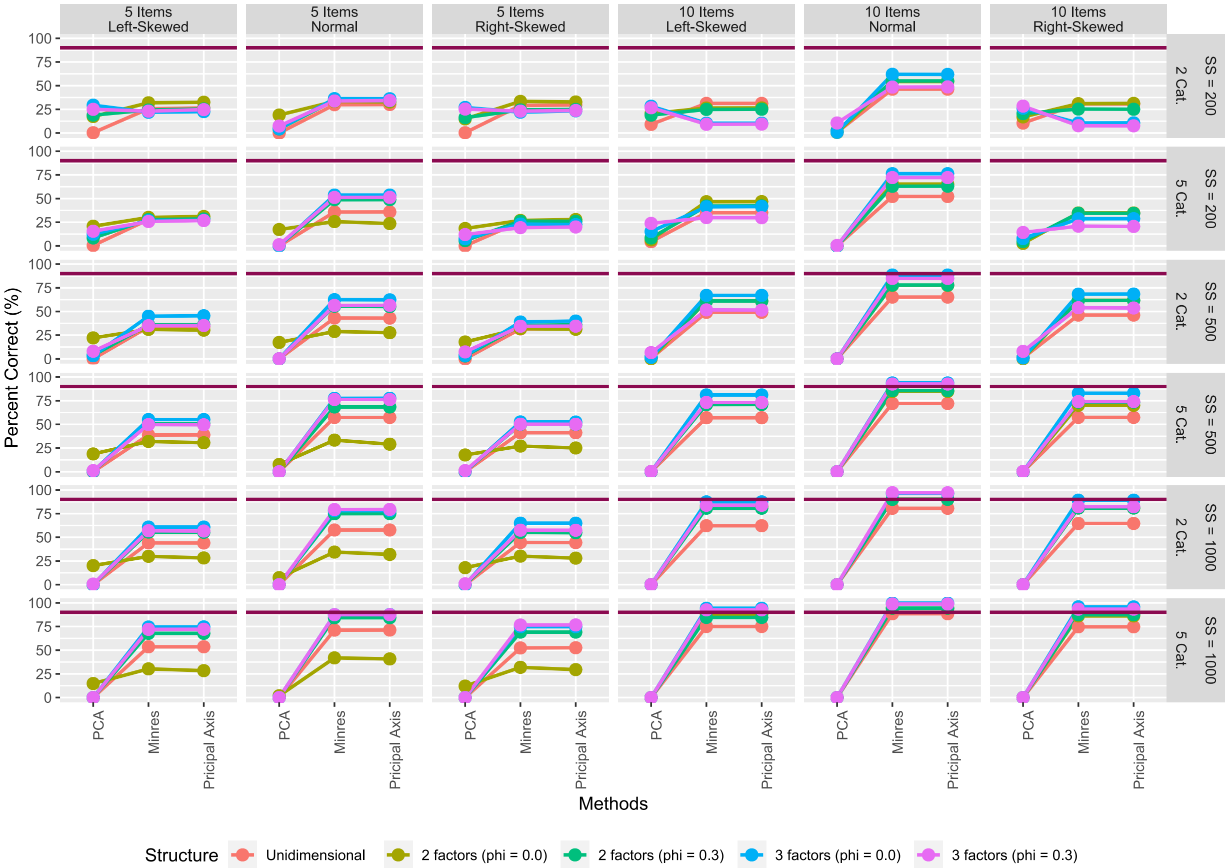
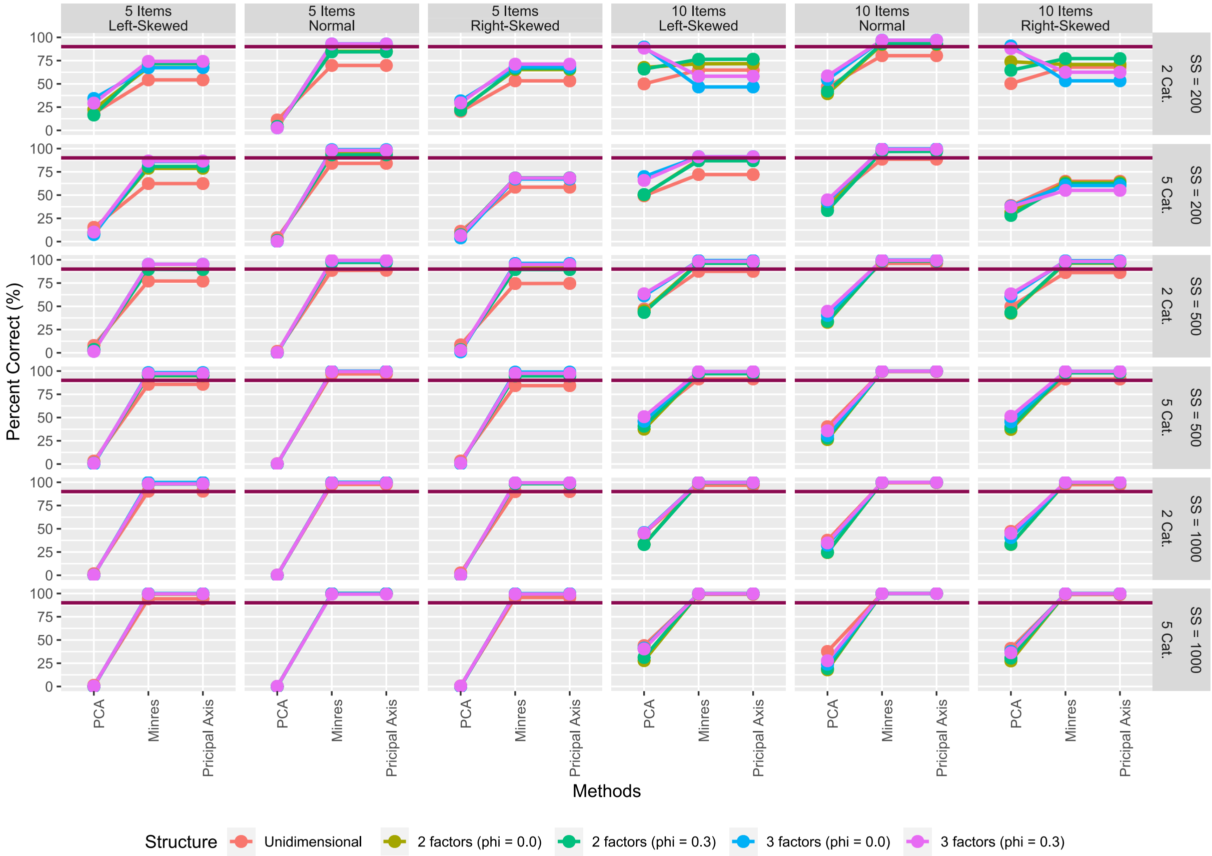


Figure 5. *PC values of methods for average factor loading is 0.70.*



4. DISCUSSION and CONCLUSION

We compared PCA, MINRES, and PAF methods regarding bias and percent correct values. As a result of the study, all methods gave unbiased results in conditions with high average factor loading. However, when evaluated in terms of percent correct, PCA performed adequately in none of the conditions, and the other methods performed adequately in about 30%. Although this result may seem contradictory, it is related to the calculation of the RB and PC values. For example, in the simulation condition where the average factor loading is 0.40, the average factor loading is estimated to be 0.43 in all 1000 replications. In this case, the average of 1000 replications will be obtained as 0.43. The RB value will be calculated as 0.08 ($\frac{0.43-0.40}{0.40}$ = 0.075) and will be considered biased within the acceptable range. However, in terms of PC, since not all 1000 replicates are in the range of 0.38-0.42 (the average factor loading is 0.43 in all replicates), the PC value will be 0. This indicates that even if unbiased estimates are made, there are inaccurate estimates in terms of accuracy. Therefore, a decision can be made by evaluating the methods in terms of both bias and accuracy.

The methods were overestimated in almost all PCA conditions when the average factor loading was low. This result is consistent with De Winter and Dodou (2016). However, MINRES and PAF gave unbiased results in almost all conditions where the average factor loading was low, and underestimation was observed in a small part of the conditions. According to this, the fact that the variables are skewed, the number of categories or the measured construct is multidimensional does not affect the performance of MINRES and PAF much. Since there is an underestimate in the already biased results, it can be considered as the lower limit of the factor loadings obtained when MINRES or PAF is used in EFA. For this reason, it can be said that similar quality results will be obtained in similar samples due to its use in scale development studies.

According to studies that compare PCA and PAF like Snook and Gorsuch (1989) and Widaman (1993), PAF outperformed PCA in most of the conditions, especially for shorter tests. PCA overestimated loadings across all factors. Differences between estimated and population loadings have decreased if loadings get higher. Our study is consistent with Snook and Gorsuch (1989) and Widaman's (1993) study with this line. We found that if the factor loadings get higher, RB values get lower.

From this point of view, the preference of PCA in scale development studies, especially in cases where the average factor loading is low, may cause the scale to appear of higher quality than it is. When this situation affects reproducibility, it may cause the scale to give different results from the results in the development study, even if it is used in similar samples. For this reason, attention should be paid to whether PCA is used in scale development studies, and the results should be evaluated with this sensitivity.

When the results are evaluated in terms of PC values, in simulation conditions with low factor loadings, PCA did not perform adequately in any condition. In contrast, MINRES and PAF performed adequately in approximately 10% of the conditions. In conditions with low factor loadings, skewness of distribution, the number of categories, and items per factor affect the performance of these methods. However, PC values can be considered a more conservative statistic since they show what percentage of all replications are within ±5% of the actual factor loading.

When the simulation conditions affecting the RB and PC values are analyzed, it is observed that the method used ($\eta^2 = 0.60$) and average factor loading ($\eta^2 = 0.10$) affect the RB values more. It can be said that items per factor ($\eta^2 = 0.16$), average factor loadings ($\eta^2 = 0.48$), sample size ($\eta^2 = 0.15$), and methods ($\eta^2 = 0.75$) are effective on PC values. The mean and median of RB and PC values for all conditions are similar (~80%).

4.1. Recommendations

As a result of this study, researchers who develop or adapt scales may be advised not to use PCA when using EFA as a factor extraction method. If they use PCA, factor loadings should be taken into consideration, as they are mostly overestimated. In the current literature review, PCA is still the most commonly used factor extraction method (see Table 1). However, it should be considered that the factor loadings obtained from these scales are overestimated. Researchers who will use the developed scales should consider this when selecting scales. It can be said that the reported factor loadings can be considered as the upper limit of the actual factor loadings. In addition, this situation will create problems in terms of both reliability and reproducibility. Therefore, for skewed, two- or five-category data, it may be recommended that practitioners use the MINRES or PAF method regardless of the number of dimensions and correlations between dimensions.

Researchers may conduct similar simulation studies on variables with different numbers of categories or mixed-format data. In future studies, comparing PCA with other methods in terms of inter-factor correlation may be considered.

4.2. Limitations

In this study, smoothing was performed when calculating the tetrachoric correlation matrix, especially in the case of skewed distribution of two-category data sets. Therefore, the results obtained should be evaluated within the framework of smoothing bias. However, considering that smoothing will also be required in real data sets with skewed distributions, it can be said that the results will be similar to the real situations. In addition, in this study, categorical data was handled with only 2 and 5 categories. It should be taken into account that in real situations, different numbers of categories (such as 3, 4, 6, or 7) may be encountered. It would not be appropriate to generalize these results to all categorical data. In addition, in the simulation study, thresholds were used to categorize the variables. This causes each variable to have a different skewness coefficient. Although the average skewness coefficient is ±2.5, it should be taken into consideration that not all variables have this value but have values close to it. The study, the k/p ratio is considered as the items per factor ratio, 5/1 and 10/1. There is a need for studies with higher items per factor.

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 authors.

Contribution of Authors

Tugay Kaçak: Investigation, Systematic review, Writing-original draft, and Formal analysis.
Abdullah Faruk Kılıç: Methodology, Supervision, Coding for simulations, and Writing-original draft.

Orcid

Tugay Kaçak https://orcid.org/0000-0002-5319-7148
Abdullah Faruk Kılıç https://orcid.org/0000-0003-3129-1763

REFERENCES

Beauducel, A., & Herzberg, P.Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(2), 186-203. https://doi.org/10.1207/s15328007sem1302_2

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.

Collins, L.M., Schafer, J.L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330–351. https://doi.org/10.1037/1082-989X.6.4.330

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

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*(1), 1–9. https://doi.org/10.7275/JYJ1-4868

Coughlin, K. (2013). An analysis of factor extraction strategies: A comparison of the relative strengths of principal axis, ordinary least squares, and maximum likelihood in research contexts that include both categorical and continuous variables. https://www.semanticscholar.org/paper/7bbfc6050a24dce49454a56d9af2ad2e6fc40ad2

De Winter, J.C.F., & Dodou, D. (2016). Common factor analysis versus principal component analysis: A comparison of loadings by means of simulations. *Communications in Statistics - Simulation and Computation, 45*(1), 299-321. https://doi.org/10.1080/03610918.2013.862274

Fabrigar, L.R., & Wegener, D.T. (2012). *Exploratory factor analysis*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001

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

Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Forero, C.G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 625-641. https://doi.org/10.1080/10705510903203573

Garson, G.D. (2023). *Factor analysis and dimension reduction in R: A social scientist's toolkit*. Routledge, Taylor & Francis Group.

Ho, A.D., & Yu, C.C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement, 75*(3), 365–388. https://doi.org/10.1177/0013164414548576

Howard, M.C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction, 32*, 51-62. https://doi.org/10.1080/10447318.2015.1087664

Jöreskog, K.G. (2003). *Factor analysis by MINRES*. https://www.ssicentral.com/wp-content/uploads/2021/04/lis_minres.pdf

Kaplan, D. (Ed.). (2004). *The SAGE handbook of quantitative methodology for the social sciences*. SAGE.

Kılıç, A.F., & Doğan, N. (2021). Comparison of confirmatory factor analysis estimation methods on mixed-format data. *International Journal of Assessment Tools in Education, 8*(1), 21–37. https://doi.org/10.21449/ijate.782351

Koyuncu, İ., & Kılıç, A.F. (2019). The use of exploratory and confirmatory factor analyses: A document analysis. *Eğitim ve Bilim, 44*(198), 361-388. https://doi.org/10.15390/EB.2019.7665

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Lozano, L.M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology, 4*(2), 73–79. https://doi.org/10.1027/1614-2241.4.2.73

Mabel, O.A., & Olayemi, O.S. (2020). A comparison of principal component analysis, maximum likelihood and the principal axis in factor analysis. *American Journal of Mathematics and Statistics, 2*(10), 44–54.

MacCallum, R.C., & Tucker, L.R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin, 109*(3), Article 3. https://doi.org/10.1037/0033-2909.109.3.502

Oranje, A. (2003, April 21). Comparison of estimation methods in factor analysis with categorized variables: Applications to NEAP data [Paper presentation]. *Annual Meeting of the National Council on Measurement in Education*.

Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research* (Version 2.4.3) [Computer software]. Northwestern University. https://cran.r-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Sigal, M.J., & Chalmers, R.P. (2016). Play it again: Teaching statistics with Monte Carlo simulation. *Journal of Statistics Education, 24*(3), 136-156. https://doi.org/10.1080/10691898.2016.1246953

Snook, S., & Gorsuch, R. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin, 106*, 148-154. https://doi.org/10.1037/0033-2909.106.1.148

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

Watkins, M.W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219–246. https://doi.org/10.1177/0095798418771807

West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Sage.

Widaman, K. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28*(3), 263–311. https://doi.org/10.1207/s15327906mbr2803_1

Zygmont, C., & Smith,M.R. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. *The Quantitative Methods for Psychology, 10*(1), 40-55. https://doi.org/10.20982/tqmp.10.1.p040

APPENDIX

Appendix 1. *Thresholds.*

Categories	Right Skewed (S.C. = 2.5)	Normally Distributed	Left Skewed (S.C. = -2.5)
2	$Y = \begin{cases} 0, y_i \leq 1.178 \\ 1, y_i > 1.178 \end{cases}$	$Y = \begin{cases} 0, y_i \leq 0.00 \\ 1, y_i > 0.00 \end{cases}$	$Y = \begin{cases} 0, y_i \leq -1.178 \\ 1, y_i > -1.178 \end{cases}$
5	$Y = \begin{cases} 0, & y_i \leq 1 \\ 1, & 1 < y_i \leq 1.189 \\ 2, & 1.189 < y_i \leq 1.5 \\ 3, & 1.5 < y_i \leq 2.1 \\ 4, & y_i > 2.1 \end{cases}$	$Y = \begin{cases} 0, & y_i \leq 1{,}5 \\ 1, & -1{,}5 < y_i \leq -0{,}5 \\ 2, & -0.5 < y_i \leq 0.5 \\ 3, & 0{,}5 < y_i \leq 1{,}5 \\ 4, & y_i > 15 \end{cases}$	$Y = \begin{cases} 0, & y_i \leq -2 \\ 1, & -2 < y_i \leq -1.7 \\ 2, & -1.7 < y_i \leq -1.2 \\ 3, & -1.2 < y_i \leq -0.99 \\ 4, & y_i > -0.99 \end{cases}$