Article

A Comparison of Methods for Determining the Number of Factors to Retain in Exploratory Factor Analysis for Categorical Indicator Variables

Holmes Finch

Department of Educational Psychology, Ball State University, Muncie, IN 47306, USA; whfinch@bsu.edu

Abstract: Exploratory factor analysis (EFA) is a widely used tool in the social sciences. Researchers employ it to identify the latent structure underlying observed indicator variables during the process of scale development, theory construction, and comparison of various constructs. One of the most important aspects of conducting EFA is determining the number of factors to retain. There exist a number of techniques for this purpose, but none have been identified as uniformly optimal in all situations. The purpose of this simulation study is to compare several such techniques in the context of dichotomous and ordinal indicator variables (corresponding to items on an instrument). Some of the methods investigated in this study include well-established techniques, such as parallel analysis and the minimum average partial correlation, as well as newly developed ones, such as out-of-sample prediction error and the next eigenvalue sufficiency test. The results of the study demonstrate that a Bayesian estimation approach and the out-of-sample prediction error method are particularly effective for identifying the number of factors to retain. The implications for practice are discussed.

Keywords: factor analysis; parallel analysis; network modeling; out-of-sample error

1. Introduction

Exploratory factor analysis (EFA) is perhaps one of the most widely used statistical methods in psychology and other social sciences. For example, EFA plays a role in scale development, helping researchers to identify the latent structure underlying measurement instruments (Ratti et al., 2017). In other contexts, psychologists and educational researchers use EFA to investigate theories about constructs of interest, such as motivation or executive functioning (Coker et al., 2018). Finally, it can be used as a precursor to confirmatory factor analysis (CFA; Canivez et al., 2019). One of the major challenges in using EFA is determining how many factors to retain. This issue arises as a result of the inherently exploratory nature of EFA. In contrast to CFA, in which the hypothesized factor structure is explicitly modeled by the researcher, for EFA, the number and nature of the underlying factors is not specified a priori. Therefore, the researcher must use the results from the EFA to ascertain how many latent variables are likely to underlay the observed indicators.

There are a number of techniques available to researchers for determining the number of factors to retain from an EFA, none of which has been found to be the best in all conditions. Prior work with EFA has found that methods that are effective with normally distributed indicator variables may not perform as well for non-normal or categorical indicators (e.g., Yang & Xia, 2015; Wirth & Edwards, 2007). The goal of the current study is to compare two relatively new techniques for determining the number of factors to retain (the next

eigenvalue sufficiency test and out-of-sample prediction) with other approaches that have been shown to be effective in prior work. Prior work in this area, particularly with the newer methods, has largely focused on continuous observed indicator variables. This study extends upon this earlier work by using categorical indicators. The methods selected for inclusion in this study were selected based upon their proven performance in prior research (e.g., parallel analysis, minimum average partial, and exploratory graph analysis), or because they are new, and have shown promise in prior work, but have not been studied in the context of categorical indicator variables (e.g., next eigenvalue sufficiency test, out-of-sample prediction error, Bayesian EFA). Prior research and the relative merits of these methods are described in more detail below. In the following pages there is a brief review of EFA, followed by a description of the various methods used in this study. The study goals are then outlined, the simulation methodology used to address these goals is described, and the results of the simulation are presented. Finally, the manuscript concludes with a discussion of the results and implications for practice.

1.1. Exploratory Factor Analysis

The standard EFA model can be expressed as follows:

$$Y = v + \Lambda \xi + \Psi \tag{1}$$

where

Y = matrix of observed indicator variables
ξ = matrix of factor(s)
v = vector of intercepts
Λ = matrix of factor loadings relating indicators to factor(s)
Ψ = matrix of unique random errors associated with the observed indicators

Of particular interest in the context of EFA is Λ, which links the observed indicator variables to the factors. Two of the most popular approaches for estimating the parameters in Equation (1) are maximum likelihood (ML) and principal axis factoring (PAF). Once the factor loadings have been estimated, they are rotated, with the goal of achieving a simple structure such that each variable has only a single large loading.

Perhaps the most important step in conducting EFA is determining the number of factors to retain. EFA is inherently exploratory in nature, and the parameters in Equation (1) are freely estimated such that the indicators are allowed to have non-zero loadings for any (or all) of the latent variables. Furthermore, the number of factors in EFA is not defined ahead of time. Researchers typically try several EFA solutions, each differing by the number of factors to be retained. Including additional factors will always yield a better statistical fit to the data, in much the same way that more complex regression models generally account for more variance in the dependent variable (Gorsuch, 1983). Therefore, if the researcher's only goal is to more fully explain the covariance matrix for the observed variables, then including more factors will always be the preferred strategy. However, in most research scenarios, we want to explain the observed data with the most parsimonious model possible. In the EFA context, this means retaining the fewest number of factors possible, while still providing satisfactory statistical fit to the observed variance–covariance matrix. In addition, the EFA solution should yield a conceptual meaningful solution. Balancing statistical fit with conceptual coherence can be a difficult task. Thus, quantitative researchers have devoted much attention to the development of statistical tools to help researchers to identify the number of factors to retain in an EFA. Below is a description of several of these that have proven to be effective in a number of situations, and which will be examined in the current study.

1.2. Parallel Analysis

One of the most consistently accurate methods that has been developed for determining the number of factors to retain is parallel analysis (PA), which was first described by Horn (1965). PA is based upon the use of bootstrap sampling to create a large number of datasets that conform to the null hypothesis of 0 factors. For each of these datasets, EFA is conducted, and the eigenvalues associated with each factor are retained. The eigenvalues from the observed data are compared to those from the bootstrap sample data in order to determine the number of factors to retain. If the observed eigenvalue for a factor is greater than the 95th percentile for that factor in the bootstrap sample, the null hypothesis is rejected and the factor is retained. As an example, if the observed eigenvalue associated with the first factor is greater than the 95th percentile for the first factor in the bootstrap sample, then the researcher would determine that at least one factor should be retained. The observed eigenvalue for the second factor would then be compared to the 95th percentile of the second factor for the bootstrap sample, and again, the factor would be retained if the observed eigenvalue is greater than the bootstrap 95th percentile. This process is repeated until the null hypothesis is not rejected. Previous research has demonstrated that PA is a very effective tool for determining the number of factors to retain (Auerswald & Moshagen, 2019; Fabrigar & Wegener, 2011; Preacher & MacCallum, 2003; Xia, 2021).

1.3. Minimum Average Partial

The minimum average partial (MAP) approach to determining the number of factors to retain was developed by Velicer (1976). It uses the correlations among the observed indicators, after the effects of the factors are partialed out, to determine the number of factors to retain. The correlation matrix for the observed indicators is first calculated, and the correlations are squared and averaged. Next, a principal components analysis (PCA) is fit to the data. The correlation matrix among the observed variables is again calculated, this time after the influence of the first factors is partialed out. Again, the correlations are squared and averaged. This step is repeated for each factor solution of interest (e.g., 2, 3, 4). The researcher would then retain the number of factors that corresponds to the minimum of the average partial correlations obtained using the steps described above. Researchers have found MAP to be an accurate approach for determining the number of factors that should be retained in an EFA (Caron, 2018; Garrido et al., 2011; Ruscio & Roche, 2012; Zwick & Velicer, 1986).

1.4. Comparison of Model Fit Statistics

Some researchers have treated the determination of the number of factors to retain as a model fit problem, such that the factor model that yields the best fit to the data is deemed to be optimal (Clark & Bowles, 2018; Garrido et al., 2011; Preacher et al., 2013). One commonly used set of fit indices for model selection in statistics is information indices such as the Bayesian Information Criterion (BIC). The BIC takes the following form:

$$BIC = \chi_M^2 + q\,ln(n)v \tag{2}$$

where

χ_M^2 = model Chi-square value
q = number of parameters estimated in the model
v = number of observed variables
n = sample size

Statistics such as the BIC reflect the overall fit of a model to the data, with a penalty for model complexity. Models with more factors will always yield a better fit to the data,

and thus will need to be penalized for their relatively great complexity when compared to simpler models. In other words, models with more factors need to be penalized for their relatively great complexity so that they are not selected simply because they are more complex. This is the logic behind information indices such as the BIC. Models with lower BIC values are taken to provide a better fit to the data. Thus, in the context of EFA, the factor model with the fewest factors is the one to be retained.

In addition to using information indices as a way to determine the number of factors to retain, other scholars have suggested the use of fit indices such as the root mean squared error of approximation (RMSEA). This statistic is used to assess the fit of a factor analysis model, measuring the degree to which the model Chi-square differs from the model degrees of freedom. When the model provides a perfect fit to the data, the Chi-square and degrees of freedom will be equal to one another. Larger values of the RMSEA indicate a greater difference between the Chi-square and degrees of freedom, and thus a worse model fit. The RMSEA takes the following form:

$$RMSEA = \sqrt{\frac{\chi_T^2 - df_T}{df_T(n-1)}} \tag{3}$$

where

χ_T^2 = ML-based Chi-square test for the target model, i.e., the model of interest
df_T = degrees of freedom for the target model (number of observed covariances and variances minus number of parameters to be estimated)
n = sample size

In order to compare model fit and determine the number of factors to retain from an EFA, the difference between the RMSEA values for solutions with an adjacent number of factors (i.e., one vs. two, two vs. three, etc.) was used. Prior research (e.g., Finch, 2020; Barendse et al., 2015; Yang & Xia, 2015; Preacher et al., 2013) has found that this RMSEA difference approach can yield accurate results for both continuous and categorical indicator variables. Based on this earlier work, an RMSEA difference of 0.015 or greater suggests a meaningful difference in model fit. In other words, when the difference in the RMSEA values for two factor models (e.g., three and four factors) exceeds 0.015, the model with the smaller RMSEA is taken to provide a better fit. If the difference does not exceed 0.015, then the simpler model (i.e., the one with fewer factors) is retained.

1.5. Exploratory Graph Analysis

Recently, authors have proposed the use of an alternative network-based approach to modeling data that have traditionally been addressed using latent variable models, such as factor analysis (Epskamp et al., 2017; H. F. Golino & Epskamp, 2017; H. Golino et al., 2020). Exploratory graph analysis (EGA) uses a Gaussian graphical model (GGM; H. F. Golino & Epskamp, 2017). The GGM corresponds directly to the factor model (H. Golino et al., 2020). The inverse of the covariance matrix for the indicator variables used in factor analysis is at the heart of GGM estimation. In this context, this inverse is called the precision matrix and is denoted as K, with diagonal elements (k_{jj}, k_{ii}). The standardized negative elements of K can be used to obtain the partial correlation between the pairs of indicators *i* and *j*, as shown in Equation (4).

$$\rho_{ij} = -\frac{k_{ij}}{\sqrt{k_{jj}}\sqrt{k_{ii}}} \tag{4}$$

These partial correlations serve as the degree of relationship between pairs of variables, and are used to identify variable clusters in EGA. Variable clusters correspond to factors in an EFA.

A network reflecting the system of relationships among the variables is then estimated using the graphical least absolute shrinkage and selection operator (GLASSO) technique (Friedman et al., 2008), which is a regularization method based on the lasso estimator (Tibshirani, 1996). This approach applies penalties to the relationships among the variables, so that small values are driven to 0 and only large estimates remain. When estimating relationships among observed indicators using GLASSO, this penalized approach results in a sparse network, with only the most salient connections being estimated and all others being set to 0. The amount of regularization is determined by a tuning parameter known as g. H. Golino et al. (2020) showed that the optimal value of g can be determined using the extended Bayesian Information Criterion (eBIC). The interested reader is referred to Golino et al. for more details.

An alternative approach to fitting graphs for the purpose of modeling connections among observed indicators is the Triangulated Maximally Filtered Graph (TMFG) approach (Massara et al., 2016). This technique limits the number of zero-order correlations that can be included in a network, thereby constructing a relatively sparse network, where only nontrivial connections among the indicators are included. The number of variable clusters (corresponding to latent variables) to retain in the context of EGA is determined using the Walktrap algorithm (Pons & Latapy, 2005), details of which can be found in H. F. Golino and Epskamp (2017).

H. F. Golino and Epskamp (2017) reported that EGA yielded comparable results to PA in terms of correctly identifying the number of factors to retain. In addition, EGA outperformed MAP in this regard across a range of study conditions. EGA was also found to perform similarly to PA in the presence of outliers among the observed indicator variables (Finch, 2019). In short, EGA appears to be a viable alternative for researchers to use in determining the number of latent traits to retain when the indicators are continuous variables. However, less work has been carried out examining its performance in the context of categorical indicators. For these reasons, it is included in the current study.

1.6. Next Eigenvalue Sufficiency Test (NEST)

One of the earliest approaches for determining the number of factors to retain was described by Kaiser (1970). This technique involved retaining factors associated with eigenvalues greater than 1, based on the logic that when standardized, each observed indicator variable will have a variance of 1. Thus, if a factor accounts for more than one unit of variance (as represented by an eigenvalue greater than 1), then it is worthy of retention. A major problem with using eigenvalues as outlined above is that the cut-off of 1 has been shown not to be very accurate in many situations, in large part due to the fact that it does not account for the presence of sampling variability in the eigenvalues (Preacher et al., 2013). In order to address this problem, Achim (2017) introduced the next eigenvalue sufficiency test (NEST) as a way to more accurately assess the number of factors to retain in EFA.

The NEST involves a series of hypothesis tests, where the null is that *k* factors should be retained. The NEST sequence uses the following steps:

1. Conduct a principal components analysis (PCA) for the observed data and retain the eigenvalues.
2. Start with H_0 : retain zero factors.
3. Generate data with zero underlying factors.
4. Conduct a PCA on the generated data from step 3.
5. Compare the first observed eigenvalue from step 1 with the first eigenvalues from step 4.

6. If the observed first eigenvalue is ≥95th percentile of the generated first eigenvalues, reject H_0. Otherwise, retain H_0 and stop.
7. Generate data with one underlying factor.
8. Conduct a PCA on the generated data from step 7.
9. Compare a first observed eigenvalue from step 1 with the first eigenvalues from step 8.
10. If the observed second eigenvalue is ≥95th percentile of the generated second eigenvalues, reject H_0. Otherwise, retain H_0 and stop.
11. Continue incrementing the eigenvalues until H_0 is not rejected.

Research has demonstrated that the NEST is a promising technique for determining the number of factors to retain when the indicators are continuous variables (Brandenburg & Papenberg, 2024). It has not been thoroughly explored for use with categorical indicators, which is one of the goals of this study.

1.7. Out-of-Sample Prediction Error

Haslbeck and van Bork (2024) recommended an approach to determining the number of factors based on the ability of a given factor model to accurately predict values for the indicator variables. This approach uses the predicted covariance matrix (Σ) among the observed indicator variables for a factor model with k latent variables to make predictions of individual values for the indicators. The factor model with the most accurate predictions of the observed variables is selected as optimal. Haslbeck and van Bork took advantage of the correspondence between standardized regression coefficients, relating the observed indicators with one another and the inverse of Σ(K).

$$\frac{\beta_{i,j} + \beta_{j,i}}{2} = -\frac{K_{ij}}{\sqrt{K_{ii}}\sqrt{K_{jj}}} \tag{5}$$

where

$\beta_{i,j}$ = standardized coefficient relating indicator *j* to indicator *i*
K_{ij} = element *i,j* in the inverse of Σ

For each indicator variable x_i a regression model using all other indicators as predictors is fit to the data, where the regression coefficients come from the factor model using Equation (5). The resulting equation is used to predict values of x_i for each member of the sample.

In order to avoid the problem of overfitting that is inherent when we use the same sample to estimate the model and assess its accuracy, Haslbeck and van Bork used *k*-fold cross-validation. This approach involves creating *k* (e.g., 10) subsamples of the data, estimating the factor model with the individuals not in subsample *k*, and then obtaining predicted indicator variable values for those in subsample *k*. The prediction accuracy is then estimated for subsample *k*. These steps are repeated for each of the *k* subsamples, and the prediction errors are then averaged. The factor model that yields the lowest average prediction is error is selected.

The work of Haslbeck and van Bork (2024) is based on an approach developed by Browne and Cudeck (1989). In this earlier iteration of the method, the sample is divided in half, and a set of factor models is fit to the first half of the dataset. The resulting factor model is then used to obtain estimated indicator covariance matrices for the data in the second half of the sample. For each factor model, the resulting predicted covariance matrix for the second half of the sample is compared with the observed covariance matrix, and the optimal factor model is determined to be the one that minimizes the difference between the model predicted and observed covariance matrices. As with the Haslbeck and van Bork

approach, the retained number of factors corresponds to the solution with the smallest average prediction error for the covariance matrix.

1.8. Bayesian EFA

The final approach to be considered in this study involves Bayesian estimation of the EFA model (BEFA). The methodology underlying this approach was outlined by Conti et al. (2014), and involves the familiar Markov Chain Monte Carlo (MCMC) approach to Bayesian estimation. We will not describe Bayesian estimation in detail here. The interested reader is referred to available descriptions of Bayesian estimation, such as Kaplan (2014). In the context of EFA, MCMC is used to estimate each parameter in Equation (1), as well as the optimal number of factors to retain, K. Unlike with most standard EFA estimation procedures (e.g., ML, PAF), with BEFA, each indicator is constrained to load on only one latent trait. In other words, BEFA does not allow indicators to have cross-loadings. Essentially, as described by Conti et al., the BEFA algorithm identifies the statistically optimal arrangement of loadings relating the indicators to the factors, under the constraint that for each indicator, there can be only one non-zero loading. All the possible arrangements of the loading space are explored, and the set that best replicates the observed covariance matrix is selected as optimal. The default noninformative prior distributions for the model parameters as described by Conti et al. are as follows:

$$\Lambda : \text{N}\left(0, \sigma_m^2\right)$$
$$\xi : N(0, \Omega)$$
$$\sigma_\varepsilon^2 : IG(c,\ C)$$
$$\text{K} : \text{Dirichlet}(k)$$

A summary of the approach for determining the number of factors to retain for each method appears in Table 1.

Table 1. Determination of number of factors to retain by method.

Method	Determination of Number of Factors
Parallel analysis	Comparison of the observed eigenvalues with the bootstrap distribution of eigenvalues under the null of no factor structure. Retain a factor if its observed eigenvalue is greater than the 95th percentile of the null distribution.
Minimum average partial	Retain the number of factors corresponding to the minimum average correlation among indicators after partialing out the variance due to the factors.
Comparison of model fit statistics	Use the difference in the RMSEA values to determine the number of factors to retain. When the difference in the RMSEA between factor solutions exceeds 0.015, retain the model with the smaller RMSEA.
GLASSO	The number of factors to retain is a penalized parameter that is estimated by the model using the Walktrap algorithm. More specifically, it is the value that minimizes the penalized least squares criterion.
Next eigenvalue sufficiency test	A comparison of the observed eigenvalues from the principal components of the observed data with the eigenvalues for the data generated under the null hypothesis of *m* (e.g., 0, 1, 2…) factors. If the observed eigenvalue for *m* + 1 factors is greater than the 95th percentile of the eigenvalues for the *m* + 1 factor, given the data generated assuming a latent structure of *m* factors, then reject the null of *m* factors, and generate data assuming *m* + 1 factors. Continue until the null hypothesis is no longer rejected.
Out-of-sample prediction error	Split the sample into a training and test set. Fit an EFA model for *m* factors using the training set. Use the factor model obtained from the training set to predict the values of the indicators for the test set. Repeat for all the desired values of *m* factors and retain the number of factors that minimizes the average prediction error for the indicator variables.
Bayesian EFA	Fit the EFA model and estimate the number of factors to retain as a model parameter. The number of factors to retain is the median of the posterior distribution for the number of factors parameter.

1.9. Study Goals

The goal of this simulation study is to extend upon earlier work by comparing the performance of several methods for determining the number of factors to retain when the observed indicator variables are categorical. Prior research has found that PA and EGA, in particular, perform well under a number of conditions. However, there are new methods for determining the number of factors to retain, out-of-sample and NEST, that have not been extensively studied for models involving categorical indicators. In addition, relatively little work has been carried out examining the performance of Bayesian EFA in the context of categorical factor indicators. Thus, the current study seeks to extend the literature by ascertaining how EFA based on this estimation method compares to the more traditional methods for determining the number of factors to retain. Given that in many cases, EFA is applied to tests and assessments involving categorical item responses, it is important to know how accurate these various approaches are for such indicators.

Based upon the prior research discussed above, it is hypothesized that PA and GLASSO will yield more accurate results regarding the number of factors to retain than the other techniques, including model fit statistics and MAP. Given the lack of research involving categorical data with the out-of-sample and NEST methods, it is not clear how they are likely to perform compared to the other methods studied here. However, it has been shown that both methods are effective in the context of normally distributed indicators (Haslbeck & van Bork, 2024; Achim, 2017), so if these trends hold for categorical data, these methods should yield accuracy rates comparable to those of PA and GLASSO. In addition, it is hypothesized that in conjunction with the robust correlation matrix estimation approach, PA will be the most accurate technique for identifying the number of factors to retain.

2. Materials and Methods

The study goals described above were addressed using a Monte Carlo simulation design. For each combination of conditions, which are described below, 1000 replications were generated. The data generation and analyses were all conducted using R version 4.3 (R Development Team, 2024). A variety of factors were manipulated, each of which is described below. Data were generated using the MonteCarloSEM package (version 0.0.8) in R (Orcan & Imam, 2024). The manipulated study conditions are described below.

2.1. Number of Factors

Simulations were conducted in which the number of factors in the data-generating model were either 1, 3, or 5. These conditions were included in order to represent a range of possibilities, from a simple unidimensional model to a complex model with 5 factors.

2.2. Factor Loading Values

Three factor loading conditions were included in this study: 0.5, 0.65, and 0.8. These values were drawn from prior research (Liu & Zumbo, 2012; Shi et al., 2020; Finch, 2020; Xia, 2021), and represent relatively small, medium, and large relationships between the observed indicators and latent variables.

2.3. Number of Indicators per Factor

For each of the factors in the 3 and 5 factors cases, either 3, 6, or 12 indicators were simulated, yielding a total number of 9, 18, or 36 indicators in the 3 factors condition, and 15, 30, or 60 indicators in the 5 factors condition. For the 1 factor case, 6 or 12 indicators were used.

2.4. Indicator Categories

Indicators were simulated to be either ordinal with 4 categories, or dichotomous. When the indicators had 4 categories, the indicator threshold parameters were −0.5, 0, and 0.5, whereas for the dichotomous variables, the thresholds were 0 for all indicators. The numbers of 2 and 4 categories per indicator were selected for this study for two reasons. First, these are relatively commonly seen numbers in practice, and thus the results of the study should be useful to researchers and practitioners. Second, we can directly compare the impact of doubling the number of response categories (from 2 to 4), which may help to give instrument developers insights regarding the impact of having more versus fewer categories for their items. We recognize that other numbers of item response categories are quite common in the literature, particularly 5. We elected to include two number of category conditions in the current study in order to keep the presentation of the results at a manageable level. However, we acknowledge that other viable options exist for the number of indicator categories, and therefore we encourage future research examining these.

2.5. Interfactor Correlation

The interfactor correlations were simulated to be either 0, 0.4, or 0.8, reflecting small/no, medium, and large relationships among the latent variables (Cohen, 1988). A similar range of values has been used in previous research investigating the performance of exploratory factor analysis (Liu & Zumbo, 2012; Shi et al., 2020; Xia, 2021; Haslbeck & van Bork, 2024).

2.6. Sample Size

The sample sizes used in the study were 250, 500, and 1000. These values correspond to those used in Liu and Zumbo (2012), and represent samples ranging from relatively small (250) to relatively large (1000) in the context of EFA practice.

2.7. Methods for Determining the Number of Factors

In order to determine the number of factors to be retained, multiple methods were used for each simulation replication, including PA, MAP, EGA using GLASSO, EGA using TMFG, RMSEA, BIC, out-of-sample prediction of the covariance matrix (COV), out-of-sample prediction of the observed variable values (VAR), Bayesian EFA with informative priors (Bayes-I), Bayesian EFA with noninformative priors (Bayes-N), and NEST. PA and MAP were conducted using the R psych package, version 2.4.3 (Revelle, 2022), and EGA was conducted using the EGAnet R package, version 1.2.3 (Hudson, 2024). BEFA was employed using the befa function from the BayesFM R package, version 0.1.5 (Patek, 2024). For BEFA-N, the noninformative priors outlined above were used, whereas for BEFA-I, an informative Dirichlet prior was used, such that a probability of 0.75 was associated with the correct number of factors to retain. A maximum of 6 factors were considered by both BEFA-I and BEFA-N. For both BEFA approaches, a total of 11,000 links were used in the MCMC chain, with the first 1000 serving as the burn-in period. To address the potential of autocorrelation in the MCMC estimates, every 10th element in the chain was sampled, yielding a posterior distribution of 1000 values for each model parameter. The MAP, BIC, and RMSEA values were obtained using the vss function from the psych package.

With respect to PA, principal axis factoring was used to extract the factors, and the 95th percentile of the simulated eigenvalues was used as the cut-off against which the observed eigenvalues were compared. The PA comparison data were generated using resampling of the sampled observations, and 1000 datasets were used. Principal axis factoring was applied to each of these randomly sorted datasets, and the resulting eigenvalues were retained to create the comparison distribution. For each simulated dataset, each method

was applied and the resulting number of factors to retain was recorded. These were then used to calculate the study outcomes, which are described below. The simulation codes are available at the Open Science Foundation (https://osf.io/w6ra3/, accessed on 25 November 2024).

2.8. Study Outcomes

Two study outcomes were used in the study. The primary outcome was the proportion of replications for each combination of conditions that correctly identified the number of factors to retain. The second outcome was the mean number of factors that were recommended for retention across replications. In order to identify the manipulated results and their interactions that impacted the primary study outcome, a mixed effects analysis of variance (ANOVA) was used, in conjunction with the partial η^2 effect size. The within-subjects effect was the method used to determine the number of factors to retain, and the between-subjects effects included the other manipulated factors. For each combination of conditions, the proportion of correct outcomes across the 1000 replications was calculated, and then served as the dependent variable for the ANOVA. The assumptions of normality and heteroscedasticity were assessed and found to be met. The ANOVA was conducted using SPSS version 29.

3. Results

3.1. Three and Five Factors

The results of the ANOVA for the three/five factor simulation results identified the interactions of method by number of loadings by interfactor correlation ($F_{40,488} = 4.11$, $p < 0.001$, $\eta^2 = 0.25$), method by number of indicators by number of factors ($F_{10,119} = 16.91$, $p < 0.001$, $\eta^2 = 0.59$), and method by sample size ($F_{20,240} = 3.70$, $p < 0.001$, $\eta^2 = 0.24$) as statistically significantly associated with the proportion correct. There were no statistically significant results for the number of indicator categories, nor any interactions involving this variable. Therefore, it will not be discussed further in this manuscript.

Figure 1 displays the proportion of replications for which the correct number of factors was identified by factor loading value, interfactor correlation, and method. Regardless of the interfactor correlation, the correct identification rates were higher for models with higher factor loading values. The correct identification rates were also higher for lower interclass correlation values. In other words, the more differentiated the factors were in the population (corresponding to lower correlations), the more accurate all of the methods were in identifying the number of factors to retain. With regard to comparisons across the methods, COV, VAR, PA, and Bayes (both informative and noninformative) yielded the highest accuracy rates when the interfactor correlation was 0.8. It should be noted that except for factor loading values of 0.8, these accuracy rates were approximately 0.4 or lower. When the interactor correlation was 0.4, GLASSO consistently yielded the most accurate results for loadings of 0.65 or lower, followed by COV, VAR, and the Bayesian methods. When the loadings were 0.8, several methods had accuracy rates above 0.9, including COV, VAR, BIC, RMSEA, GLASSO, and TMFG. Finally, when the factors were completely uncorrelated, COV, VAR, GLASSO, TMFG, and the Bayes methods yielded the most accurate results for factor loadings of 0.5 and 0.65. When the loadings were 0.8, these methods, along with BIC, RMSEA, and PA, had accuracy rates between 0.95 and 1.0.
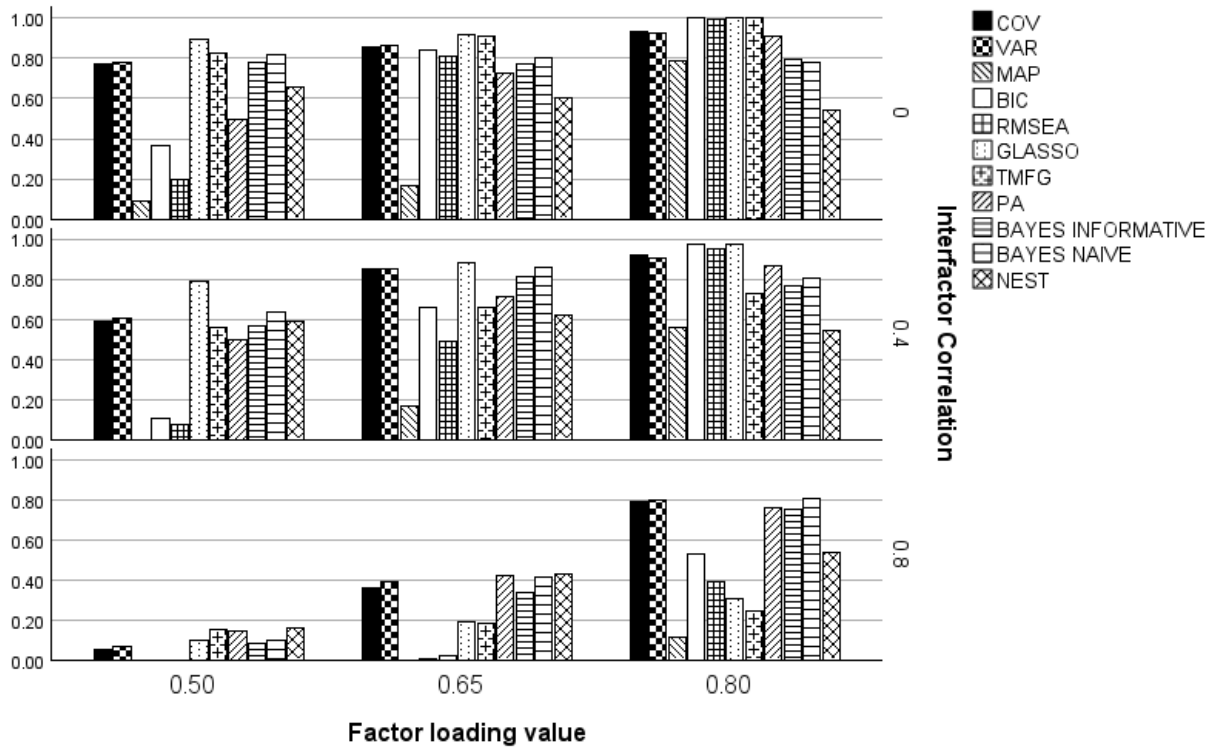
Figure 1. The proportion of replications for which the correct number of factors was identified by method, factor loading value, and interfactor correlation: three and five factors.

Figure 2 shows the proportion of replications for which the correct number of factors was selected by method, number of indicators, and number of factors. As is evident in Figure 1, COV and VAR consistently had relatively high accuracy rates. In addition, GLASSO and TMFG displayed among the highest accuracy rates when six indicators per factor were present. On the other hand, in the three indicators and three factors condition, the highest accuracy rates were associated with PA and the Bayes approaches. For three indicators and five factors, in addition to COV and VAR, PA and NEST had the highest accuracy rates among the methods examined here. The accuracy rates by method and sample size appear in Figure 3. For most of the methods, accuracy improved concomitantly with increased sample sizes. This improvement was more marked for some methods, including COV, VAR, BIC, and PA. For the other techniques, improved accuracy with increasing sample sizes was much less notable.

The results of the ANOVA for the three/five factor simulation results identified the interaction of method by number of factors by interfactor correlation ($F_{40,488} = 5.34$, $p < 0.001$, $\eta^2 = 0.29$), method by factor loadings by number of factors ($F_{10,119} = 7.11$, $p < 0.001$, $\eta^2 = 0.35$), number of indicators by number of factors ($F_{15,119} = 4.88$, $p < 0.001$, $\eta^2 = 0.22$), and method by sample size by number of factors ($F_{20,240} = 6.56$, $p < 0.001$, $\eta^2 = 0.27$) as statistically significantly associated with the proportion correct. Figure 4 includes the mean number of factors retained by method, number of factors, and interfactor correlation. Most of the techniques included in this study had means quite close to the actual number of factors used to generate the data. Generally speaking, the Bayes approaches retained too many factors when they were incorrect, whereas MAP, BIC, and RMSEA tended to underfactor when they were incorrect. Similar patterns for the mean number of factors retained are apparent in Figures 5–7, which display this metric by method and factor loading values, number of indicators, and sample size, respectively.
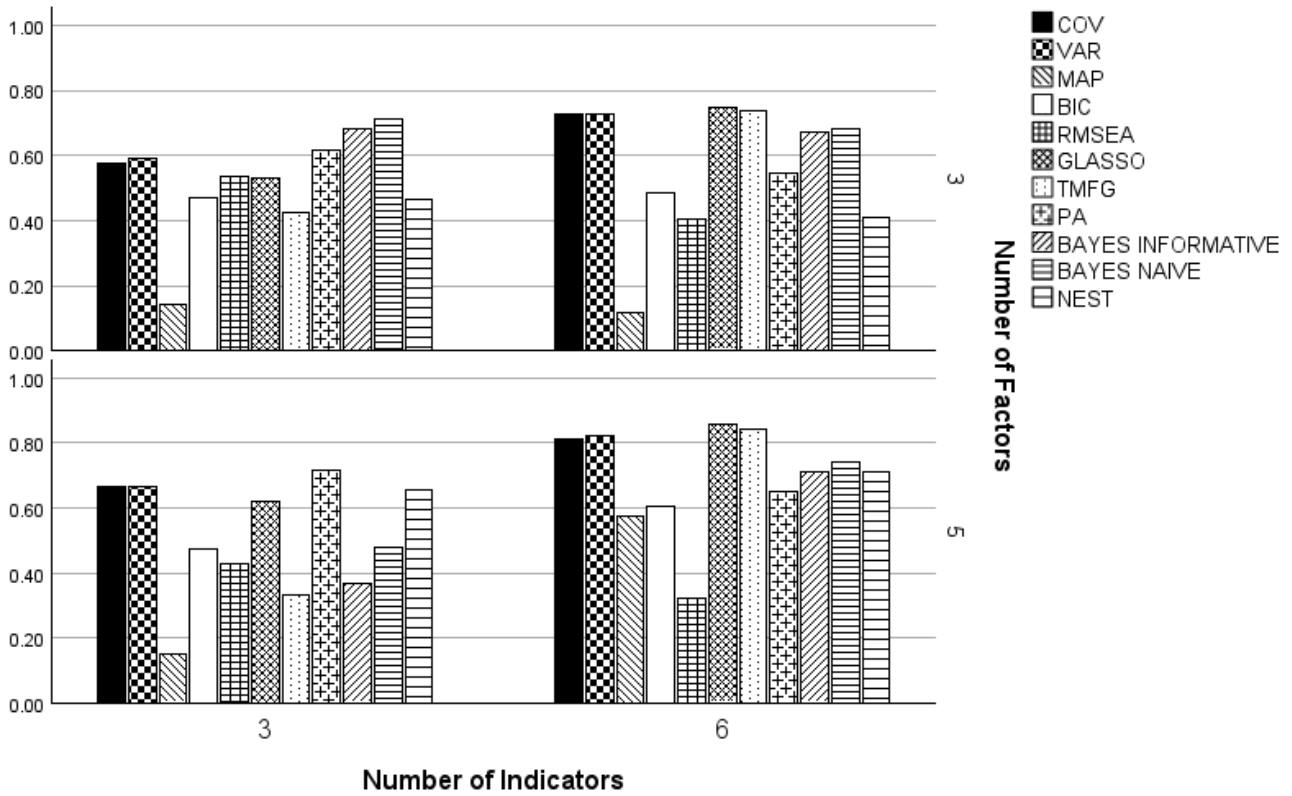
Figure 2. The proportion of replications for which the correct number of factors was identified by method, number of indicators, and number of factors: three and five factors.
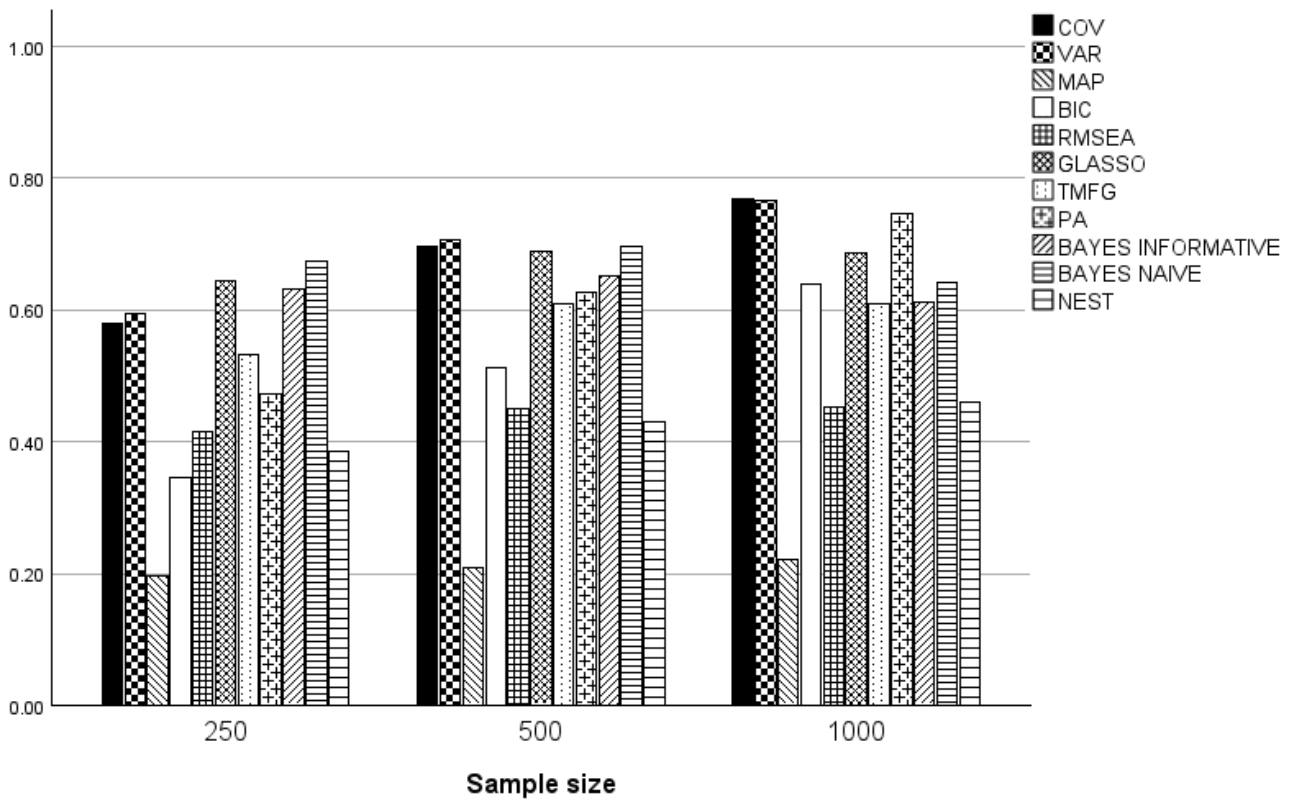

Figure 3. The proportion of replications for which the correct number of factors was identified by method and sample size: three and five factors.
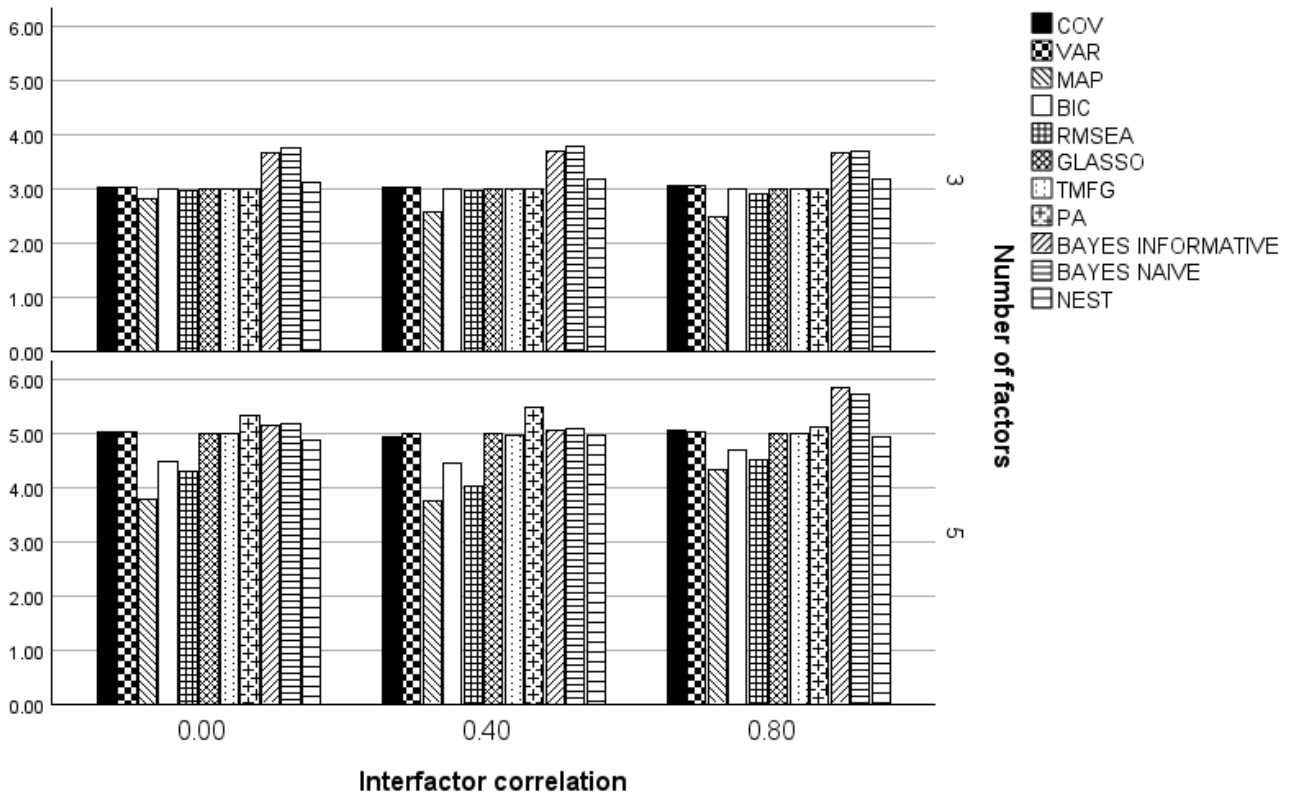
Figure 4. Number of factors retained by method, number of factors, and interfactor correlation: three and five factors.
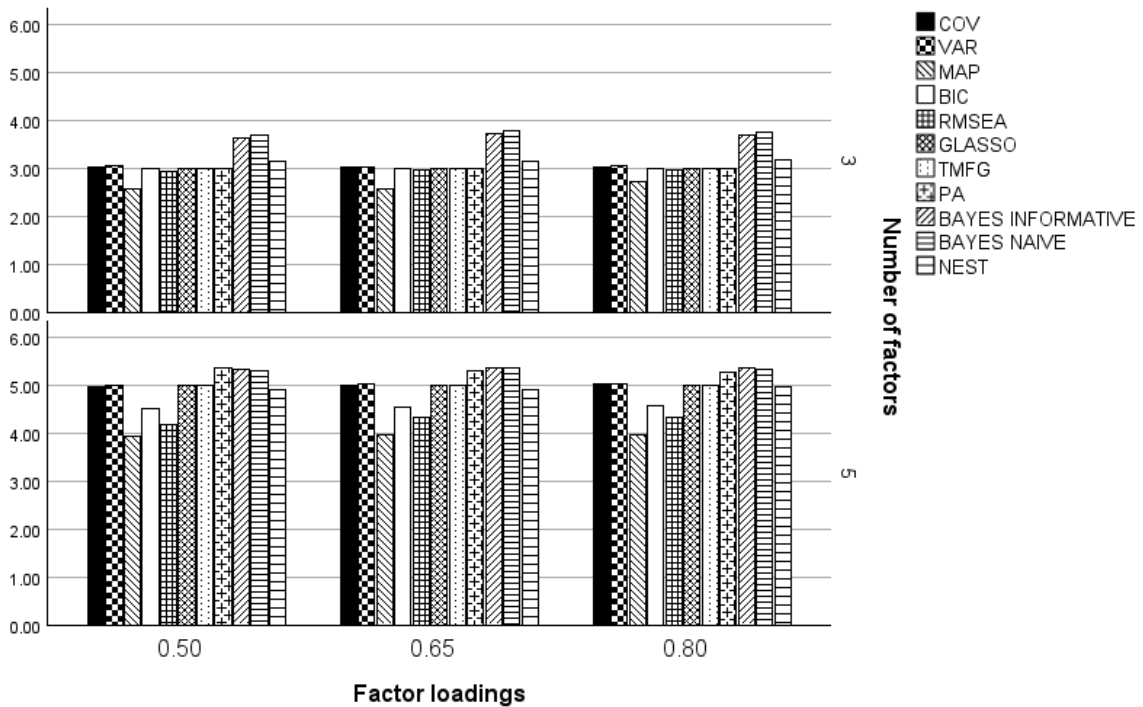

Figure 5. Number of factors retained by method, number of factors, and factor loading values: three and five factors.
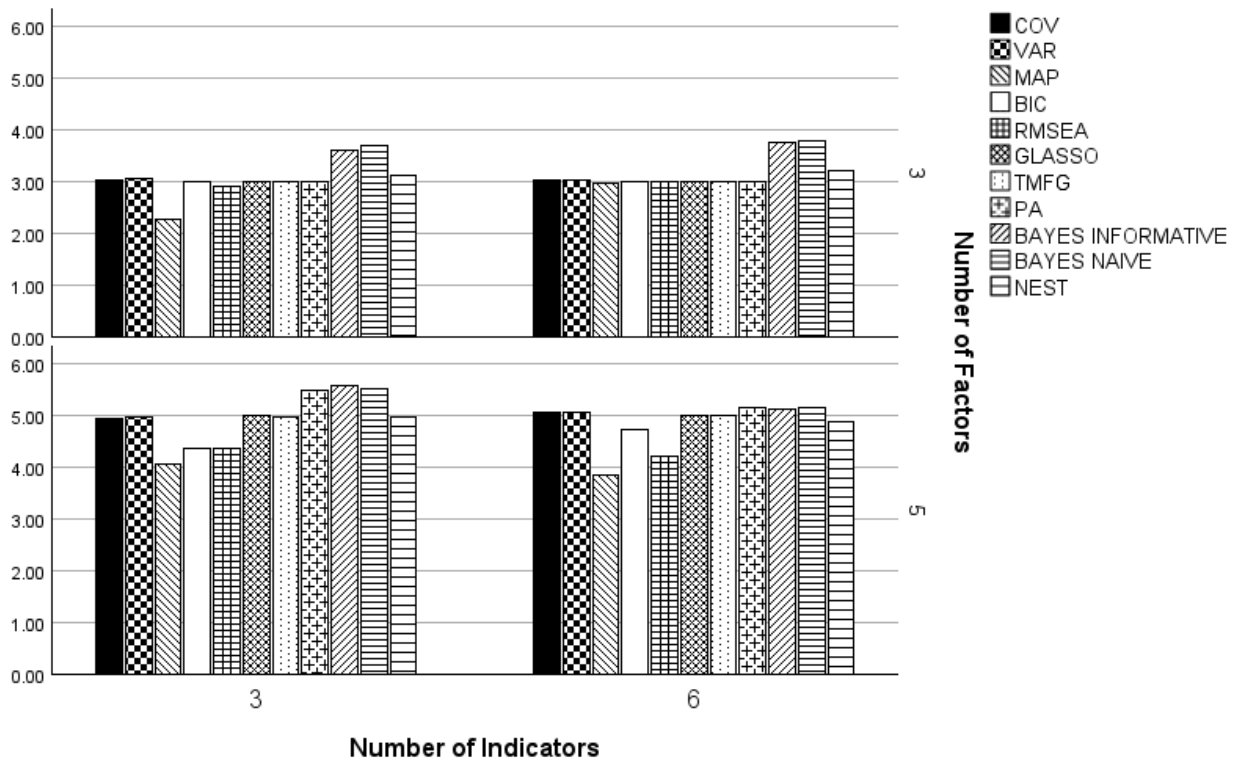
Figure 6. Number of factors retained by method, number of factors, and number of indicators: three and five factors.
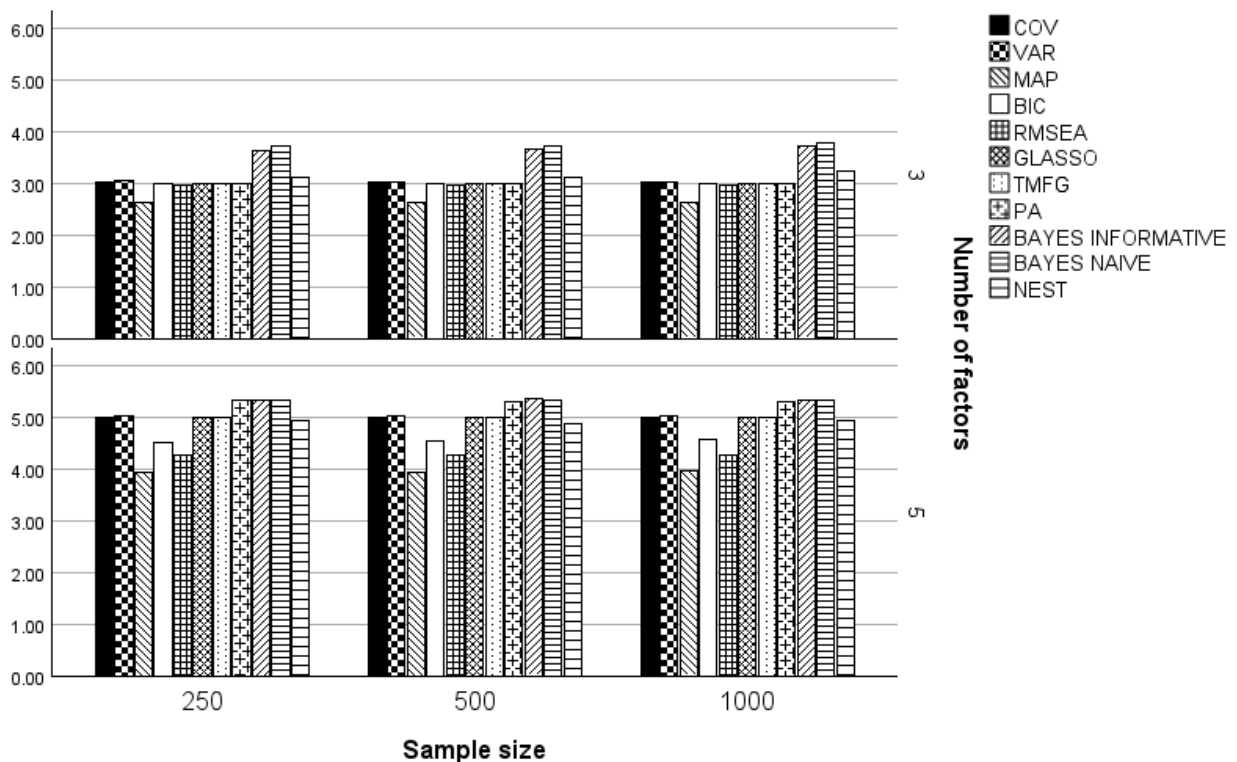

Figure 7. Number of factors retained by method, number of factors, and sample size: three and five factors.

3.2. One Factor

For the one factor case, the results of the ANOVA identified the interaction of method by sample size ($F_{20,40} = 1.93$, $p = 0.038$, $\eta^2 = 0.49$) and method by number of indicators

$(F_{10,40} = 13.10, p < 0.001, \eta^2 = 0.77)$ as statistically significantly associated with the proportion correct for factor retention. Figure 8 includes the proportion correct for the one factor case by sample size and method. Several methods were accurate in virtually all instances, including MAP, BIC, and the Bayesian estimators. In addition, for sample sizes of 500 or 1000, RMSEA, GLASSO, TMFG, and PA also had accuracy rates at or near 1.00. On the other hand, unlike RMSEA, GLASSO, TMFG, and PA, COV, VAR, and NEST each had lower accuracy rates than the other methods across sample sizes. In addition, these techniques did not see improved accuracy concomitantly with increased sample size in the one factor case.
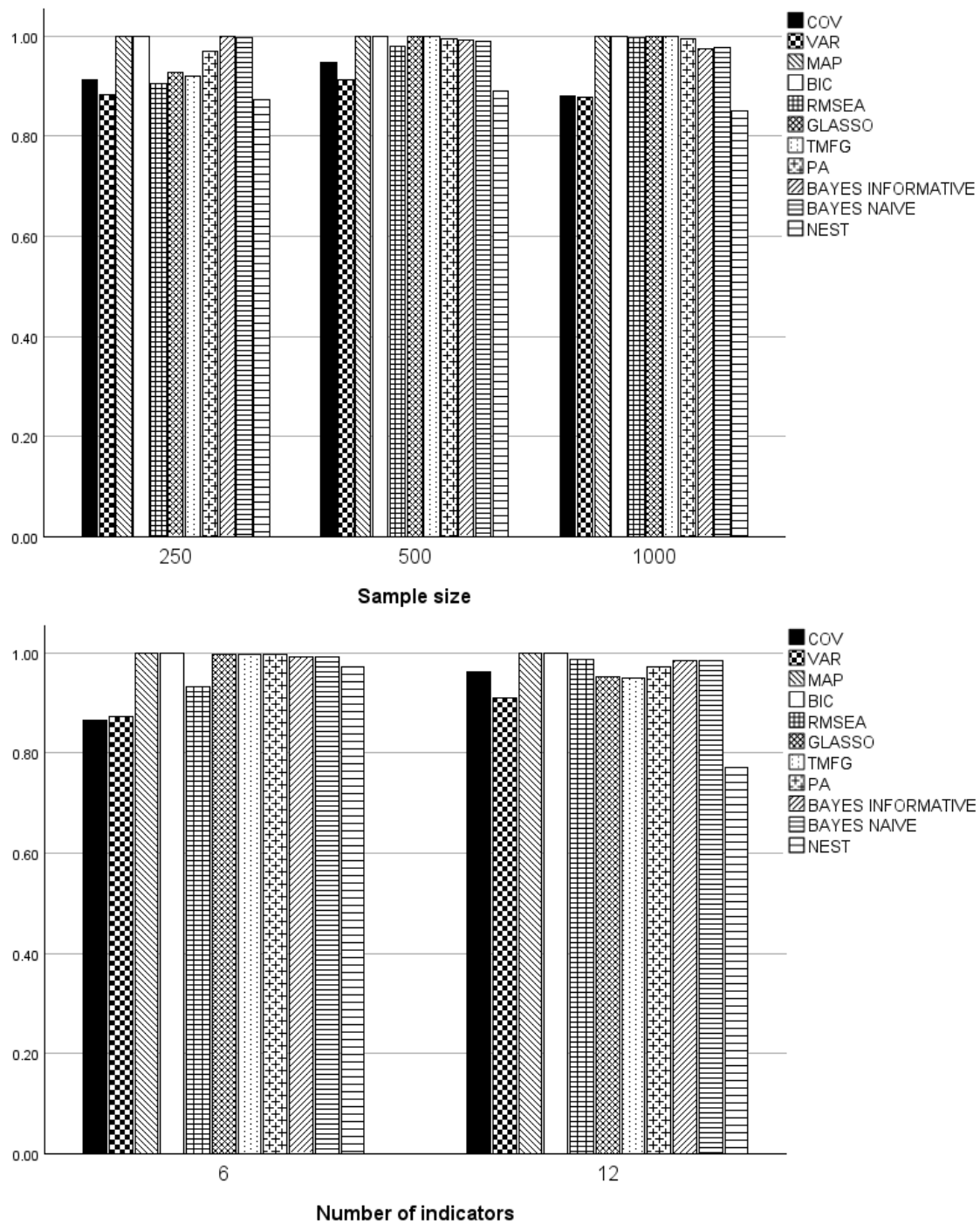

Figure 8. The proportion of replications for which the correct number of factors was identified by method, sample size, and number of indicators: one factor.

The proportion correct in the one factor case by method and number of indicators also appears in Figure 8. Consistently with the results by sample size, MAP, BIC, and the Bayesian methods all yielded accuracy rates at or near 1.0, regardless of the number of indicators. COV and VAR saw improved accuracy when more indicators were present, whereas GLASSO, TMFG, and particularly NEST all displayed somewhat lower accuracy for more indicators.

The mean number of factors retained by method, number of indicators, and sample size appear in Table 2. These results reinforce the findings from Figure 8 that MAP and BIC were extremely accurate in terms of correctly identifying the one factor solution. When COV, VAR, PA, and Bayes erred, they retained too many factors. Indeed, this overfactoring was consistent across the number of indicators and sample size conditions. In contrast, the RMSEA slightly underfactored the solutions, given that the mean number retained fell between 0.90 and 0.98. GLASSO, TMFG, and NEST appear to have been most strongly impacted by the number of indicators and sample size. In the presence of 12 indicators, these techniques retained a higher mean number of factors than was the case for 6 indicators. With respect to sample size, the means for GLASSO and TMFG were close to or at the nominal number of factors (one) for sample sizes of 500 or 1000. The mean number retained was also lower for NEST, though it was still higher than for the other methods studied here.

Table 2. Mean number of factors retained by method, sample size, and number of indicators: one factor.

Method	Number of Indicators		Sample Size		
	6	12	250	500	1000
COV	1.06	1.09	1.08	1.07	1.07
VAR	1.04	1.09	1.07	1.06	1.06
MAP	1	1	1	1	1
BIC	1	1	1	1	1
RMSEA	0.98	0.90	0.93	0.95	0.96
GLASSO	1.03	1.10	1.13	1.02	1
TMFG	1.04	1.12	1.15	1.03	1
PA	1.02	1.14	1.10	1.07	1.07
Bayes-I	1.03	1.07	1.04	1.05	1.06
Bayes-N	1.04	1.07	1.05	1.06	1.07
NEST	1.06	1.87	1.55	1.46	1.44

4. Discussion

The primary goal of this study was to compare the performance of several methods with respect to determining the number of factors to retain in the context of EFA. Prior research in this area has found that EGA and PA are consistently effective in this regard, particularly for normally distributed indicator variables (H. F. Golino & Epskamp, 2017). Recently, new approaches for determining the number of factors to retain have been suggested as viable alternatives to these other approaches, specifically the out-of-sample (Haslbeck & van Bork, 2024) and NEST (Achim, 2017) methods. Furthermore, BEFA (Conti et al., 2014) has also emerged as a potentially useful tool for researchers to use in fitting EFA models. Although these latter three approaches have been studied in the context of continuous indicators, they have not been as thoroughly examined for cases where the indicators are categorical in nature. Thus, it was hoped that the current study would extend the literature in determining the number of factors to retain by focusing on EFA models with categorical indicators, which correspond to situations where the variables of interest are items on a psychological or educational instrument.

The results of this study revealed that several approaches performed well in several conditions. Indeed, across simulation conditions, both the Bayesian methods were consis-

tently among the most accurate with respect to the number of factors to retain. The use of informative and noninformative priors did not seem to impact the performance of BEFA in any of the conditions studied here. As has been found in earlier research (REFERENCES), GLASSO and TMFG were also quite accurate when the interfactor correlation was 0 or 0.4 and the loadings were 0.65 or 0.8. EGA performed less well than several of the other methods when the interfactor correlation was 0.8. In contrast, COV, VAR, PA, and the BEFA methods were the most accurate in the highest interfactor correlation condition. It should again be stated, however, that for loadings of 0.50 or 0.65, even these methods had relatively low accuracy rates for the number of factors to retain, with values of 0.4 or less. The MAP, BIC, and RMSEA were, generally speaking, less accurate than the other methods studied here, except for the one factor case.

4.1. Implications for Practice

The results presented above provide several implications for researchers who need to determine the number of factors to retain from an EFA applied to categorical indicators. First, when the quality of the factors is low (i.e., small factor loadings) and factor separation is low (i.e., large interfactor correlations), no method studied here will provide accurate results with respect to the number of factors to retain. Second, in the other simulated cases, GLASSO, BEFA, COV, and VAR were generally among the best-performing methods. This does not mean that any one of these approaches uniformly yielded the most accurate results in all cases, but they were consistently among the best performers studied here. Thus, researchers would generally be well served in using any or all of them to determine the optimal number of factors to retain in the context of categorical indicator variables. Third, building upon the prior implication, the results of this study suggest that researchers consider using multiple techniques to determine the number of factors to retain when they have categorical indicators. In other words, one might employ GLASSO, BEFA, COV, and VAR with their data and then examine the results to see if a consensus is reached by the methods. If so, then this consensus would serve as the optimal solution regarding the number of factors to retain. The fourth implication from this study is that the use of informative and noninformative priors with BEFA generally yielded the same results, meaning that researchers may not need to be particularly concerned with the choice of priors. Fifth, the number of categories in the indicators did not seem to have an impact on the performance of these methods when it came to determining the number of factors to retain. Thus, researchers can feel confident that the techniques that they choose for this purpose will yield similar rates of accuracy whether the indicators have two or four categories. Sixth, and finally, for categorical indicators, GLASSO appears to yield more accurate results regarding the number of factors to retain than TMFG does.

An alternative approach to fitting EFA models that was not considered here, but which does show promise, is regularized EFA. There exist multiple algorithms to carry out this analysis, all of which share the goal of identifying only those loadings that are clearly different from zero. One such approach, sparse estimation via nonconcave penalized likelihood in the factor analysis model (FANC), involves the use of the minimax convex penalty function (Hirose & Yamamoto, 2015). This penalty is applied in conjunction with the standard maximum likelihood estimator, and places a penalty on all factor model parameters, including loadings. The consequence of this penalty is to drive factor values down in value, meaning that small loadings will essentially be set to 0. Likewise, the FANC algorithm places a penalty on the number of factors to be retained. Another regularization algorithm for determining the number of factors is the principal orthogonal component threshold (POET), which rests on the assumption of conditional sparsity (Fan et al., 2013). This assumption asserts that, conditional on a small number of common components, the

observed indicator variables will have small covariances with one another. Of course, the GLASSO approach used in the current study is also a regularization-based approach for identifying the latent structure in the data. Comparison of the methods included in this study is an area that should be examined in future work.

4.2. Study Limitations

As with any study, the current work has limitations that future research should seek to address. First, the distribution of the observed variables was limited to two or four categories with symmetric distributions regarding the response probabilities (as seen in the symmetric thresholds). Future research should extend upon this work by including indicators with more categories, such as five or six. In addition, future work should also examine non-symmetric threshold parameters, so that the item response probabilities are skewed. Using different threshold values for the various indicators would also add to the literature in this area. The current study simulated data with a pure simple structure. However, in many real-world situations, indicators are likely to have non-zero loadings associated with more than one factor. Thus, future research should simulate cases in which some indicators have non-zero loadings with multiple factors. Finally, future research should include conditions in which the latent traits are not normally distributed as they were in this study.

5. Conclusions

Despite the limitations outlined above, the current study extends the literature regarding the number of factors to retain in the context of EFA in a number of ways. First, it examines several promising new methods (e.g., COV and VAR) in the context of categorical indicators, and finds these to be effective tools for this purpose. Second, this work reinforces earlier findings demonstrating the efficacy of using EGA to determine the number of factors. Third, this study showed that researchers should consider adding Bayesian estimation to their data analysis toolbox when it comes to fitting EFA models.

References

Achim, A. (2017). Testing the number of required dimensions in exploratory factor analysis. *The Quantitative Methods for Psychology*, *13*(1), 64–74. [CrossRef]

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468–491. [CrossRef] [PubMed]

Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Structural Equation Modeling*, *22*(1), 87–101. [CrossRef]

Brandenburg, N., & Papenberg, M. (2024). Reassessment of innovative methods to determine the number of factors: A simulation-based comparison of exploratory graph analysis and next eigenvalue sufficiency test. *Psychological Methods*, *29*(1), 21–47. [CrossRef] [PubMed]

Browne, M. W., & Cudeck, R. (1989). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258.

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale For Children—Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, *89*(2), 195–224. [CrossRef]

Caron, P.-O. (2018). Minimum average partial correlation and parallel analysis: The influence of oblique structures. *Communications in Statistics–Simulation and Computation*, *48*(7), 2110–2117. [CrossRef]

Clark, D. A., & Bowles, R. P. (2018). Model fit and item factor analysis: Overfactoring, underfactoring, and a program to guide interpretation. *Multivariate Behavioral Research*, *53*(4), 544–558. [CrossRef]

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Coker, J. L., Catlin, D., Ray-Griffith, S., Knight, B., & Stowe, Z. N. (2018). Buprenorphine medication-assisted treatment during pregnancy: An exploratory factor analysis associated with adherence. *Drug and Alcohol Dependence*, *192*, 146–149. [CrossRef]

Conti, G., Fruhwirth-Schnatter, S., Heckman, J., & Piatek, R. (2014). Bayesian exploratory factor analysis. NRN working paper. In *The austrian center for labor economics and the analysis of the welfare state.* Johannes Kepler University.

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*, 904–927. [CrossRef] [PubMed]

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.

Fan, J., Liao, Y., & Mincheva, M. (2013). Large Covariance Estimation by Thresholding Principal Components. *Journal of the Royal Statistical Society, Series B*, *75*(4), 603–680. [CrossRef] [PubMed]

Finch, W. H. (2019). *Exploratory factor analysis*. Sage.

Finch, W. H. (2020). Using fit statistic differences to determine the optimal number of factors to retain in an exploratory factor analysis. *Educational and Psychological Measurement*, *80*(2), 217–241. [CrossRef] [PubMed]

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. [CrossRef]

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's Minimum Average Partial Factor Retention Method with Categorical Variables. *Educational and Psychological Measurement*, *71*(3), 551–570. [CrossRef]

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE*, *12*(6), e0174035. [CrossRef] [PubMed]

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*(3), 292–320. [CrossRef] [PubMed]

Gorsuch, R. L. (1983). *Factor analysis*. Psychology Press.

Haslbeck, J. M. B., & van Bork, R. (2024). Estimating the number of factors in exploratory factor analysis via out-of-sample prediction errors. *Psychological Methods*, *29*(1), 48–64. [CrossRef] [PubMed]

Hirose, K., & Yamamoto, M. (2015). Sparse Estimation via Nonconcave Penalized Likelihood in Factor Analysis Model. *Statistical Computing*, *25*, 863–875. [CrossRef]

Horn, J. L. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*, *30*(2), 179–185. [CrossRef]

Hudson, G. (2024). *EGAnet: Exploratory graph analysis*. R Library.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401–415. [CrossRef]

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. Guilford Press.

Liu, Y., & Zumbo, B. D. (2012). Impact of outliers arising from unintended and unknowingly included subpopulations on the decisions about the number of factors in exploratory factor analysis. *Educational and Psychological Measurement*, *72*(3), 388–414. [CrossRef]

Massara, G. P., Di Matteo, T., & Aste, T. (2016). Network filtering for big data: Triangulated maximally filtered graph. *Journal of Complex Networks*, *5*, 161–178. [CrossRef]

Orcan, F., & Imam, K. S. (2024). *MonteCarloSEM*. R Library.

Patek, R. (2024). *BayesFM: Bayesian inference for factor modeling*. R Statistical Library.

Pons, P., & Latapy, M. (2005). Computing Communities in Large Networks Using Random Walks. In P. Yolum, T. Güngör, F. Gürgen, & C. Özturan (Eds.), *Computer and information sciences—ISCIS 2005. ISCIS 2005. lecture notes in computer science* (Vol. 3733). Springer.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics*, *2*, 13–43. [CrossRef]

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28–56. [CrossRef] [PubMed]

Ratti, V., Vickerstaff, V., Crabtree, J., & Hassiotis, A. (2017). An Exploratory Factor Analysis and Construct Validity of the Resident Choice Assessment Scale with Paid Carers of Adults with Intellectual Disabilities and Challenging Behavior in Community Settings. *Journal of Mental Health Research in Intellectual Disabilities*, *10*(3), 198–216. [CrossRef]

R Development Team. (2024). *R: A language and environment for statistical computing*. R Development Team.

Revelle, W. (2022). *Psych: Procedures for psychological, psychometric, and personality research*. R Library.

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment 24*, 282–292. [CrossRef]

Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 1–15. [CrossRef]

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B. Methodological*, *58*, 267–288. [CrossRef]

Velicer, W. F. (1976). Determining the Number of Components from the Matrix of Partial Correlations. *Psychometrika*, *41*(3), 321–327. [CrossRef]

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58–79. [CrossRef] [PubMed]

Xia, Y. (2021). Determining the number of factors when population models can be closely approximated by parsimonious models. *Educational and Psychological Measurement*, *81*(6), 1143–1171. [CrossRef] [PubMed]

Yang, Y., & Xia, Y. (2015). On the number of factors to retain in exploratory factor analysis for ordered categorical data. *Behavior Research Methods*, *47*(3), 756–772. [CrossRef]

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432–442. [CrossRef]