

Article

# A Comparison of Methods for Determining the Number of Factors to Retain with Exploratory Factor Analysis of Dichotomous Data

W. Holmes Finch

Department of Educational Psychology, Ball State University, Muncie, IN47306, USA; whfinch@bsu.edu

**Abstract:** Exploratory factor analysis (EFA) is a very widely used statistical procedure in the social and behavioral sciences. This technique features in validity studies, as well as investigations of latent structure underlying observed measurements. A primary aspect of using EFA is determining the number of factors to retain. In addition to theoretical considerations, a variety of statistical tools have been developed and recommended for use in assisting researchers with respect to factor retention. Some research has been conducted to investigate the accuracy of these methods in the case of continuous factor indicators. The purpose of the current simulation study was to extend this earlier work to situations in which the indicator variables are dichotomous, as with questionnaire or test items. Results of this study revealed that an approach based on the combined results of the empirical Kaiser criterion, comparative data, and Hull methods, as well as Gorsuch's CNG scree plot test by itself, all yielded accurate results with respect to the number of factors to retain. Implications for practice are discussed.

**Keywords:** exploratory factor analysis; factor number; dichotomous data



**Citation:** Finch, W.H. A Comparison of Methods for Determining the Number of Factors to Retain with Exploratory Factor Analysis of Dichotomous Data. *Psych* **2023**, *5*, 1004–1018. <https://doi.org/10.3390/psych5030067>

Academic Editor: Okan Bulut

Received: 9 August 2023

Revised: 5 September 2023

Accepted: 7 September 2023

Published: 13 September 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Exploratory factor analysis (EFA) is a widely used tool in the psychological sciences. For example, in a recent review of the Psycinfo search engine for the 10 years between 31 July 2013 and 31 July 2023, using the keyword “exploratory factor analysis”, a total of 560,312 separate journal articles were identified. Given its ubiquity and importance in assessing the viability of psychological theories, the appropriate application of EFA by researchers is of key importance in order to ensure that statistical results are as accurate as possible. Perhaps the central decision to be made by researchers using EFA is the number of factors to retain [1]. Many methods exist for determining the number of factors to retain when using EFA, and a number of studies have compared these methods in the presence of normally distributed observed indicator variables; e.g., [2]. However, in many psychological and educational applications, the observed indicator variables associated with the use of EFA are dichotomous in nature e.g., [3,4]. For example, popular scales of mood [5] and anxiety [6] rely on items that force respondents to select whether particular symptoms are present or absent. Likewise, personality inventories [7], academic admissions tests [8], and numerous tests of cognitive ability [9] consist of items with dichotomous responses. Therefore, it is of key importance for researchers and data analysts to have insights regarding optimal techniques for determining the number of factors to retain in the case of categorical, and in particular dichotomous, indicator variables. Although research has been conducted to examine the performance of techniques for determining the number of factors to retain in EFA, these have focused on continuous indicator variables [2]. Given the aforementioned ubiquity of dichotomous items in social science research, it is worthwhile to investigate the performance of factor determination methods for such scales.

The purpose of this study was to build on earlier work [2,10] by examining a wide range of statistics for determining the number of factors to retain in the context of dichoto-

mous observed indicator variables. The prior research in this area has focused primarily on approaches for determining the number of factors to retain when the indicator variables are continuous [2]. While certainly informative, this earlier work does not pertain directly to the case where indicators consist of dichotomous variables, such as item responses, which are common in the social sciences, as discussed above. There has been some strong recent work examining the performance of parallel analysis and revised parallel analysis in the context of EFA with dichotomous items [10]. However, this work has not included the full range of statistical tools available to determine the number of factors to retain, beyond the parallel analysis framework. The current study extends on this earlier work by comparing the full array of methods heretofore used with continuous indicators with dichotomous items.

The remainder of the manuscript is structured as follows: First is a brief review of the one-factor model, followed by a description of the various factor retention methods examined in this study. Next is a review of prior research focused on factor retention techniques with continuous and categorical indicators. The simulation methodology used in this study is then described, followed by the results, and finally a discussion of the findings from the current study in light of prior research.

## 2. Exploratory Factor Analysis

The EFA model can be expressed as:

$$Y = \Lambda\zeta + \Psi \quad (1)$$

where

$Y$  = matrix of observed indicator variables

$\zeta$  = matrix of factor(s)

$\Lambda$  = matrix of factor loadings relating indicators to factor(s)

$\Psi$  = matrix of unique random errors associated with the observed indicators

Of particular interest in the context of EFA is the factor loading matrix,  $\Lambda$ , which links the observed indicator variables to the latent factors. Indicator factor combinations with non-trivial loadings, e.g.,  $>0.3$  [11], are said to be associated with one another. Based on the indicator groupings, researchers can make inferences regarding the nature of the latent traits underlying the observed indicators. A number of methods are available for estimating the model parameters in Equation (1), a step often referred to as factor extraction. Perhaps the most widely used of these methods are maximum likelihood estimation (MLE) and principal axis factoring (PAF). After the initial factor loadings are extracted, they are typically transformed (i.e., rotated) in order to improve interpretability of the results through the mathematical encouragement of simple structure in  $\Lambda$ ; i.e., each indicator will be associated with a single factor.

## 3. Determining the Number of Factors to Retain

EFA is, by definition, an exploratory statistical technique. As such, researchers using it do not explicitly link individual indicator variables to specific latent factors. In addition, typical EFA practice involves an exploration of multiple possible factor solutions, with the optimal such approach being selected as the “winner”. A key aspect of selecting the number of factors to retain is matching the factor loading pattern with theory. The ultimate decision regarding the EFA results to retain should be conceptually coherent and theoretically sound with respect to how the indicators group together with the factors [12]. In addition, researchers can make use of several statistically based techniques for ascertaining the number of factors to retain. These methods do not refer to theory, but rather provide insights regarding the number of factors to retain based upon statistical criteria. Although these purely statistical approaches cannot be in isolation to determine the number of factors to retain, they are very useful, in conjunction with theoretical considerations, to help researchers ascertain the number of factors to retain. Following is a description of several

such techniques that have been shown in prior research to be useful tools for determining the number of factors to retain.

### 3.1. Chi-Square Test

One of the more widely used approach for determining the number of factors to retain is the chi-square statistic associated with MLE. MLE utilizes a minimization process in which model parameter estimates (i.e., factor loadings, variances, and error variances) that yield a predicted covariance matrix ( $\Sigma$ ) of the indicator variables that is very close in value to the observed covariance matrix ( $S$ ). The value of the function that results from this minimization process can be converted to a chi-square statistic that can be used to test the null hypothesis that  $\Sigma = S$ ; i.e., the model predicted covariance matrix is equal to the actual covariance matrix. This null hypothesis states that the fit of the model is perfect, which is actually rarely achieved in practice [13,14]. Therefore, many close-fitting models that prove useful may be rejected by the chi-square goodness of fit test, thereby leading a researcher to conclude that the model does not provide a good fit to the data, when in fact the proposed model does provide reasonably good data model fit. For this reason, this test may not be particularly useful for assessing the fit of an individual model [14,15].

Despite the limited utility of the MLE chi-square test for assessing the feasibility of a single model, it has been shown to be an effective tool for comparing EFA models with differing numbers of factors. This approach involves obtaining the chi-square statistic for each of several EFA results (e.g., 1 factor, 2 factors, 3 factors) and then calculating the difference in values between adjacent factor models. These differences follow the chi-square distribution if the observed indicators come from a multivariate normal distribution [16]. Thus, the chi-square difference statistic can be used to test the null hypothesis of equivalent model fit between a pair of EFA models. As an example, the researcher might fit EFA models with 1 to 4 factors. Then, the differences in chi-square fit statistic values between 1 and 2, 2 and 3, and 3 and 4 factors are calculated. If the 1 vs. 2, and 2 vs. 3 chi-square differences values are statistically significant, but the 3 vs. 4 are not, then the researcher would retain 3 factors.

This approach, commonly referred to as the sequential model test (SMT), has been shown to be an effective tool when the assumption of multivariate normality is met, and particularly when factors are relatively highly correlated [2]. However, when data do not follow a multivariate normal distribution, the population model is mis-specified, and/or minor factors are present in the population, SMT has tendency to overextract; i.e., recommend the retention of too many factors [2,17]. SMT was included in the current study in order to serve as a baseline approach, given its popularity in applied settings. However, given that the data used here were dichotomous, thus violating the assumption of multivariate normality, it was not anticipated that this approach would be particularly effective in determining the number of factors to retain.

### 3.2. Parallel Analysis

Another inferential approach for determining the number of factors to retain is parallel analysis (PA). This method was first described by Horn [18], and involves the generation of synthetic data that has the same marginal properties (i.e., means and variances) as the actual observed data, but which has no underlying latent structure; i.e., 0 factors. Most often, the synthetic data are generated nonparametrically through the random mixing of indicator variable values within variables across observations. Typically, a large number (e.g., 1000) of these synthetic datasets is generated, for each of which EFA is conducted. The eigenvalues from these analyses, which reflect the explained variance associated with each factor, are then used to create distributions of eigenvalues that would be expected if no factor structure is present. The eigenvalues obtained using the observed data are then compared to these distributions in order to determine the number of factors to retain. A factor is retained if its observed eigenvalue is larger than the 95th percentile of the distribution of null factor eigenvalues generated from the random data. For example, if the

observed first, second, and third eigenvalues were 4.5, 2.1, and 0.9 and the 95th percentiles of the first, second, and third eigenvalues were 1.4, 1.2, and 1, the researcher would retain two factors. Researchers [17] extended the use of PA based on the tetrachoric correlation matrix of dichotomous item response data and principal components analysis. Given that the data used in this study were dichotomous item responses, this approach based on the tetrachoric correlation was used in the current study. The tetrachoric correlation has been shown to be most appropriate for estimating correlation coefficients with dichotomous variables [19].

### 3.3. Revised Parallel Analysis

Researchers have pointed out that standard PA may have some limitations with respect to assessing the magnitude of all but the largest eigenvalues produced by EFA [20]. The rationale behind this argument is that it is only for the assessment of the first eigenvalue that the underlying latent structure is assumed to include no factors. If the first eigenvalue for the observed data exceeds the 95th percentile of the reference distribution, the presence of one factor can be inferred. Thus, the second observed eigenvalue should be compared with a null distribution assuming a single latent trait. Likewise, the third observed eigenvalue should be compared with a null distribution assuming two latent traits, and so on. In order to address this issue, [20] proposed the use of what they called revised PA (RPA), in which the comparison data are generated assuming that  $k - 1$  latent variables underlie the observed data, rather than 0 factors as is the case with traditional PA. Thus, when testing for 4 factors, RPA would generate data from a 3-factor model, holding the marginal distributions of the variables to be equal to those in the observed sample. In other respects, the methodology for RPA is the same as for standard PA.

### 3.4. Comparative Data Method

Concurrent with the work by [20,21] described the comparative data method (CD) for determining the number of factors to retain. CD generates a sample of 10,000 cases based on a correlation matrix associated with  $k - 1$  latent variables. Next, 500 random samples of  $n$  observations (where  $n$  is equal to the observed data sample size) are drawn from the 10,000 cases, and principal components analysis (PCA) is conducted in order to obtain eigenvalues. Inference regarding the number of factors to retain is then carried out as with PA and RPA. For PA, RPA, and CD in this study, the tetrachoric correlation was used.

### 3.5. Minimum Average Partial

Velicer [22] proposed a method for determining the number of factors to retain that is based upon an examination of the average squared partial correlations among the observed indicators, accounting for the influence of the latent variables. This method, called minimum average partial (MAP), is carried out using a multiple-step procedure. In the first step, the correlations among the observed variables are calculated, squared, and then the squares are averaged. In the second step, the squared correlations among the indicators are again calculated and averaged, in this case after partialing out the first latent variable obtained using EFA. In the third step, the average squared correlation among the observed variables is again calculated, this time partialing out the first two latent variables. These steps are repeated for the first  $p - 1$  factors, where  $p$  is the number of observed indicators. The researcher then would retain the number of factors corresponding to the minimum average squared partial correlation, as this corresponds to the point at which the maximum amount of systematic variance in the observed indicators is accounted for by the latent variables. Thus, if the average squared partial correlation values were 0.2, 0.15, 0.09, 0.03, 0.07, 0.11, and 0.12, the researcher would select the 4-factor solution. Simulation research has consistently demonstrated that MAP is one of the more accurate methods for determining the number of factors to retain [21,23–25].

### 3.6. Empirical Scree Plot Methods

A plot of the eigenvalues by the factor number, known as the scree plot, was proposed by [26] determining the optimal EFA solution. The purpose of this plot is to examine the relationship between the number of factors and the amount of variance explained in the observed variables (as measured by the eigenvalues), with an eye toward identifying for which factor number the amount of explained variance declines sharply. In practice, this is done using a scatterplot with the eigenvalue on the  $y$ -axis, the factor number on the  $x$ -axis, and a line connecting points in the plot. As noted above, the eigenvalues decrease in value from the first through the last factor, which is reflected in the scatterplot. The researcher using this approach examines the plot, looking for the point where the line connecting the eigenvalues begins to flatten out in its rate of decline. This point will correspond to the number of factors that should be retained.

In practice, the scree plot has been found to be relatively inaccurate with respect to the number of factors to retain, in large part perhaps due to the highly subjective interpretive criterion [27]. Thus, multiple more objective approaches based on the scree plot have been discussed in the literature. Gorsuch's [1] CNG scree test involves the calculation of the slope linking the first three eigenvalues, then the calculation of the slope linking eigenvalues 2, 3, and 4, then the slope linking eigenvalues 3, 4, and 5, and so on. The researcher then compares these slopes with one another, and selects the number of factors where the difference between the slopes is greatest. Thus, for example, if the largest difference between slope values lies between the line for points 2, 3, and 4 versus the line for points 3, 4, and 5, we would retain 4 factors. Zoski and Jurs [28] suggested a variant (NMREG) of the Gorsuch approach in which pairs of regression equations are estimated using all of the data points, rather than just sets of 3 at a time. Thus, for  $p$  indicator variables, the following pairs of equations would be considered:

Line 1 (eigenvalues 1, 2, and 3)	Line 2 (eigenvalues 4 through $p$ )
Line 3 (eigenvalues 1, 2, 3, and 4)	Line 4 (eigenvalues 5 through $p$ )
Line 5 (eigenvalues 1, 2, 3, 4, and 5)	Line 6 (eigenvalues 6 through $p$ )

The slopes for the lines in each pair (e.g., line 1 versus line 2) are then compared using a  $t$ -test, and the number of factors to be retained is associated with the maximum  $t$  value. As an example, if the maximum  $t$  statistic is associated with the comparison between lines 3 and 4, then 4 factors (corresponding the largest factor number in line 3) would be retained.

### 3.7. Hull Method

Lorenzo-Seva, et al. [29] described an approach for determining the number of factors to retain from an EFA that is related to the scree plot approaches described above. The Hull technique is based on model goodness of fit test results using the following steps.

1. Fit EFA models with varying numbers of factors (e.g., 1, 2, 3, etc.).
2. For each model from step 1, calculate a goodness of fit statistic ( $G_m$  and model degrees of freedom  $df_m$ ), such as the CFI.
3. Compare the values of  $G_m$  across factor models. A model is rejected if its value of  $G_m$  is lower than that of an adjacent model with fewer factors.
4. Models from those remaining after step 3 are rejected as viable if their  $G_m$  value lies below a line connecting points in a plot of  $G_m$  and  $df_m$ .
5. Repeat step 4 until no nonviable solutions are remaining.
6. Select the number of factors,  $m$ , for which the following statistic is maximized from among the set of potentially viable solutions:

$$\frac{(G_m - G_{m-1}) / (df_m - df_{m-1})}{(G_{m+1} - G_m) / (df_{m+1} - df_m)} \quad (2)$$

Prior research has shown that the Hull method using the CFI as  $G_m$  is an effective tool for identifying the number of factors with continuous indicators [2,29], particularly when it is used in combination with other statistics. This is the approach that was employed in the current study.

### 3.8. Empirical Kaiser Criterion

Braeken and van Assen [30] described a method for determining the number of factors based upon the eigenvalues, thereby distinguishing it from the Hull approach, and placing it in the broad family including the objective scree plot techniques and PA/RPA. This empirical Kaiser criterion (EKC) is an outgrowth of the familiar Kaiser criterion, in which all factors with eigenvalues greater than 1 are retained. A problem with the original Kaiser approach is that it does not account for sampling variability inherent in analyses that do not involve the entire population [27]. The EKC method is based upon an assumption that the distribution of eigenvalues follows the Marcenko-Pastur distribution [31]. The upper bound of this distribution is calculated as:

$$\lambda_{1,ref} = \left(1 + \sqrt{\frac{p}{n}}\right)^2 \quad (3)$$

where

$p$  = number of indicators

$n$  = sample size

The largest eigenvalue from the EFA (corresponding to the first factor) is compared to this maximum value. The remaining eigenvalues are compared to the variance corrected reference eigenvalues as calculated below:

$$\lambda_{m,ref} = \max\left(\frac{p - \sum_{i=0}^{m-1} \lambda_m}{p - m + 1} \left(1 + \sqrt{\frac{p}{n}}\right)^2, 1\right) \quad (4)$$

The number of factors to be retained corresponds to the number of eigenvalues that are greater than the reference value in Equation (4). For example, consider the case where the first 4 observed eigenvalues are 3.5, 2.4, 1.7, and 1.1. If there are 20 indicators and a sample size of 200, the upper bound from Equation (3) is 1.73. Thus, the first factor would be retained. If the subsequent reference eigenvalues calculated from Equation (4) were 1.5, 1.4, and 1.25, the researcher would retain a total of 3 factors.

### 3.9. Combination Approach

In addition to the individual statistics described above, a combination approach to determining the number of factors to retain was also investigated in the current study. This combination technique included a vote-counting approach to EKC, Hull, and CD, per Auerswald and Moshagen [2]. Specifically, the number of factors to retain corresponded to the largest vote count from among the methods. If multiple numbers of factors received the same number of votes, the most parsimonious model from the vote winners was selected. It is important to note that the prior authors also included SMT in their combination approach. However, given that the dichotomous data used in the current study do not conform to the multivariate normal assumption underlying MLE, the SMT approach was excluded from the combination method utilized here.

### 3.10. Limited Information Item Factor Analysis

McDonald [32] introduced the Normal-Ogive Harmonic Analysis Robust Method (NOHARM) approach to fitting factor analysis models to dichotomous item response data. The NOHARM model takes the form:

$$P(U_{ji} = 1 | \theta) = N(\beta_{j0} + \beta_{j1}\theta_{1i} + \dots + \beta_{jk}\theta_{ki}) \quad (5)$$

where

$U_{ji}$  = response to item  $j$  for respondent  $i$ ; 1 = item endorsement

$\theta$  = vector of latent traits

$\theta_{ki}$  = level on latent trait  $k$  for respondent  $i$

$\beta_{j0}$  = intercept for item  $j$

$\beta_{jk}$  = coefficient relating item  $j$  to trait  $k$

Parameter estimates for the model in Equation (5) can be obtained using a limited information approach based on unweighted least squares (ULS).

Fit of the NOHARM model to the data can be assessed using a chi-square statistic, as described by Gessaroli and De Champlain [33]:

$$\chi_{GD}^2 = (N - 3) \sum_{l=2}^J \sum_{j=1}^{J-1} z_{jl}^{(r)} \quad (6)$$

where

$N$  = total sample size

$z_{jl}^{(r)}$  = Fisher's  $z$ -transformed residual correlation,  $r$ , for item pair  $jl$

In order to determine the number of factors using  $\chi_{GD}^2$ , a sequence of models is fit to the data, differing based upon the number of factors, much as with SMT. For each such model,  $\chi_{GD}^2$  is calculated and the difference between this statistic for models with  $k$  and  $k - 1$  number of factors is calculated. This  $\Delta\chi_{GD}^2$  statistic follows an approximate  $\chi^2$  distribution [33] with the degrees of freedom equal to the difference in degrees of freedom for the two models. If the  $\Delta\chi_{GD}^2$  is statistically significant, the researcher concludes that the EFA model with  $k$  factors provides better fit than the  $k - 1$  factor. The  $k$  factor solution is then compared to the  $k + 1$  solution, and the procedure is repeated until a non-statistically significant conclusion is achieved, at which point the more parsimonious solution in the comparison (e.g.,  $k$  vs.  $k + 1$ ) is selected. For example, if the  $\Delta\chi_{GD}^2$  values comparing 1 vs. 2 and 2 vs. 3 factors were statistically significant, but the value comparing 3 vs. 4 factors was not, the researcher would retain 3 factors.

### 3.11. Full Information Item Factor Analysis

An alternative to modeling dichotomous item response data with more than one latent trait comes in the form of the full information multidimensional IRT (MIRT) model. The 2-parameter logistic (2PL) MIRT model (which corresponds to the standard factor model) for dichotomous data takes the form:

$$P(U_{ji} = 1 | \theta) = \frac{e^{(\alpha_{j1}\theta_{1j} + \dots + \alpha_{jk}\theta_{kj}) + \gamma_j}}{1 + e^{(\alpha_{j1}\theta_{1j} + \dots + \alpha_{jk}\theta_{kj}) + \gamma_j}} \quad (7)$$

where

$\alpha_{jk}$  = discrimination parameter for item  $j$  on latent trait  $k$

$\gamma_j$  = difficulty for item  $j$

Maximum likelihood can be used to obtain the parameters in Equation (7) based on the EM algorithm [34]. Alternative estimation approaches such as either a Metropolis–Hastings Robbins–Monro or Bayes can also be used in conjunction with the MIRT model. The EM algorithm was used in the current study. As with the standard EFA and NOHARM models, an exploratory implementation of the MIRT model can be used when the researcher is unsure as to the number of underlying latent traits for a given dataset. Further details regarding the implementation of the exploratory MIRT model in the R software environment can be found in Chalmers [26].

Determination of the number of factors to retain when using the full information MIRT model involves use of the likelihood ratio test, in much the same way as the SMT and  $\Delta\chi_{GD}^2$  described above. For each prospective model, the exploratory MIRT model is fit to the data and the chi-square goodness of fit statistic is obtained. As with SMT and NOHARM, the difference between these statistics is compared to the chi-square distribution with degrees of freedom equal to the difference in the degrees of freedom for the two models. An approach similar to that described above for NOHARM and the SMT can be used with exploratory MIRT in order to determine the optimal number of factors to retain. As an example, if the tests comparing 1 vs. 2 factors and 2 vs. 3 factors were statistically significant, but the test comparing 3 vs. 4 factors was not, the researcher would retain 3 factors from the EFA.

### 3.12. Prior Research Investigating Factor Retention Methods

There is a relatively large body of literature investigating the performance of multiple methods for identifying the number of factors to retain in the context of EFA. This work has been particularly focused on the case of continuous (and primarily multivariate normal) indicator variables. For example, research has demonstrated that PA and its variants are among the most effective methods for identifying the number of factors to retain in the context of EFA [35–39]. Early simulation results consistently demonstrated that PA tended to identify the correct number of factors underlying a set of observed indicators more frequently than did most other alternative approaches, such as the scree plot, the eigenvalue greater than 1 rule, and proportion of variance explained by the factors. More

recent work has shown that RPA and CDM may yield somewhat more accurate results than standard PA, and that RPA is perhaps the best performer most consistently across a variety of conditions [40]. Auerswald and Moshagen [2] found that PA based on PCA (rather than the common factor model) yielded more accurate results than did RPA based on the common factor model. Similarly, Guo and Choi [10] found that PA and RPA based on PCA was able to accurately identify the number of factors to retain when the indicators were dichotomous in nature, provided that the sample is 500 or larger. Lim and Jahng [38] conducted a simulation study comparing various versions of PA and found that standard PA yielded accurate results to within 1 factor (plus or minus) of the actual number of latent traits underlying the observed data.

With respect to the objective methods associated with the scree plot, simulation research has shown that the Zoski and Jurs approach and the CNG scree test are both very effective at determining the number of factors to retain, assuming that there are at least 3 latent variables present in the data ([27]). This work was limited to the case of multivariate normal indicator variables. Thus, the current research was designed to extend upon these earlier results by applying the objective scree plot techniques to the case of dichotomous indicator variables.

Auerswald and Moshagen [2] compared the performance of several criteria for determining the number of factors to retain when using EFA. They concluded that using results from a combination of statistics, rather than a single value, might prove to be the most effective. Based upon their simulation study, Auerswald and Moshagen recommended that when researchers fit EFA models with continuous indicator variables, the results of the SMT and either RPA, Hull, or EKC agree on the number of factors to retain, they settle on this value. However, if the SMT and the other technique do not suggest the same number of factors, then the researcher should rely primarily on CD, EKC, or RPA solely. Finally, their results demonstrated that using a combination of approaches typically yields more accurate results than does any single approach.

With respect to  $\chi^2_{GD}$ , Finch and Habing [41,42] showed that a variant of  $\Delta\chi^2_{GD}$  performed well, particularly with more factors present in the population. This approach was more effective than alternatives when the underlying latent trait was normally distributed, whereas when the trait was skewed, nonparametric alternatives were better able to control the Type I error rate [41]. Svetina and Levy [43] extended this work and found that the  $\Delta\chi^2_{GD}$  statistic was more accurate than nonparametric alternatives, particularly for data when some indicators were associated with multiple factors, the test was longer, and the factors were more highly correlated.

### 3.13. Study Goals

The primary goal of this simulation study was to compare several methods for determining the number of factors to retain in the context of EFA with dichotomous indicator variable data. This study was designed to expand on earlier work, particularly that of Guo and Choi [10] and Auerswald and Moshagen [2]. As discussed above, prior work with continuous indicator variables indicated that using a combination of methods may be the optimal approach for this purpose [2]. In addition, Guo and Choi showed that using PA based on the tetrachoric correlation matrix and PCA can provide accurate results for dichotomous item response data. With respect to specific techniques, CD, EKC, or RPA may yield the most accurate results with respect to the number of factors to retain, based on prior research. The current study builds upon this earlier work by investigating the factor retention problem when indicators are dichotomous in nature using a wide array of possible methods for this purpose. Several of the methods examined in earlier research are included in this study, as are approaches designed specifically for categorical indicator variables, including NOHARM and the MIRT model. In addition, methods that have been found to be effective for a larger number of indicators than is present in the current work (i.e., [10]) were also included in this study. Thus, the current study was designed to extend two recent studies on EFA and factor retention with a focus on dichotomous indicators and a large array of methods. This study provides direct comparisons among a wide range of methods that have not been compared in this way before. The techniques are compared with respect to the proportion of instances in which the correct number of factors is recommended by each method, as well as the mean number of factors suggested for retention.

## 4. Methods

In order to address the goals of this study, a Monte Carlo simulation study was conducted. For each combination of conditions (described below), 1000 replications were used. The simulations were carried out with the R software system, version 4.1.2 [44]. Dichotomous item response data were simulated based on a structural equation model with either 1 or 3 latent traits, using the MonteCarloSEM R package [45]. The item thresholds used in data generation were drawn from a



standardized mathematics test, with a mean of 0.04 and ranging between  $-2.3$  and  $2.7$ . The tetrachoric correlation matrix was estimated for the set of dichotomous indicators and then served as the dataset for each of the factor retention methods included in the study.

In addition to the methods used to determine the number of factors to retain, other manipulated factors included sample size, the number of underlying latent traits, interfactor correlation, the number of indicators per factor, and the factor loading value. The sample sizes used in this study were 200, 500, 1000, and 2000. These values were selected based on prior research in the area of factor analysis [2,10,46]. Data were generated either from a 1 or a 3-factor model, allowing for assessment of the factor retention methods for both unidimensional and multidimensional data. There were 5, 10, or 20 indicator variables per factor, for each of the factor number conditions, reflecting what might be seen with short subscales (5 items) to relatively long subscales (20 items). Results for the 10 and 20 indicator conditions were found to be virtually identical and therefore, in order to simplify the presentation of results somewhat, only those for 5 and 10 indicators are reported below. The factor loadings were set at either 0.6 or 0.8, which, when combined with the number of indicators, resulted in scale reliability values between 0.74 and 0.97. In addition, these loading values were selected in order to examine performance of the factor retention methods when more than 50% of the variance in each indicator was associated with the latent trait (loading = 0.8) and when less than 50% of the indicator variance was explained by the factor (loading = 0.6). Three conditions were included for interfactor correlations, including 0.25, 0.50, and 0.75. These values were selected so as to reflect small, medium, and large relationships among the factors [47].

The methods used to determine the number of factors to retain included EKC, Hull, SMT, CD, Combined, MAP, NMREG, CNG, PA, RPA, MIRT, and NOHARM. These methods were applied using the R packages 'nFactors' [48], 'EFA.dimensions' [49], 'mirt' [50], and 'sirt' [51]. The outcomes of interest were the proportion of replications for which each method correctly identified the number of factors to retain and the mean number of factors retained across the 1000 replications for each combination of conditions. In order to identify which combinations of the manipulated conditions were associated with the correct proportion of factors retained, analysis of variance (ANOVA) was used. For each combination of study conditions, the proportion of correct cases was calculated and served as the dependent variable in the ANOVA. In addition to statistical significance, the eta-squared effect size was also calculated for each term in the model. ANOVA models were fit separately for the 1 and 3-factor cases, due to the difference in the terms available for each.

## 5. Results

### 5.1. Three Factors

#### 5.1.1. Factor Retention Accuracy Rate

For the case when 3 factors were present in the population, ANOVA identified the interactions of the factor determination method, number of indicators, and interfactor correlation ( $F_{22,286} = 7.03$ ,  $\eta^2 = 0.79$ ), and the factor determination method, number of indicators, and sample size ( $F_{66,286} = 2.18$ ,  $\eta^2 = 0.33$ ) as being statistically significantly related to the proportion of replications with the correct number of factors being retained. Table 1 includes the proportion of replications with the correct number of factors by method, number of indicators, and interfactor correlations, in the 3-factors case. When there were 10 indicators per factor, EKC and Combined exhibited the highest proportion of correct number of factors retained, across interfactor correlation levels. In addition, CD, CNG and MAP had rates of correct identification of 1.00 for an interfactor correlation of 0.25 and 0.50, but rates below 0.90 when factors had a correlation of 0.75. Of these latter techniques, CD had the highest accuracy rate for the highest interfactor correlation.

**Table 1.** Proportion of replications with correct number of factors, number of indicators, and interfactor correlation: 3 factors.

I *	C	EKC	Hull	SMT	CD	Comb	MAP	NMREG	CNG	PA	RPA	MIRT	NOHARM
5	0.25	0.9951	0.6362	0.0000	0.9911	0.9951	0.5230	0.4110	1.0000	0.6986	0.5805	0.5427	0.0408
	0.50	0.8461	0.5495	0.0000	0.8911	0.8461	0.4314	0.5088	1.0000	0.6759	0.5648	0.5976	0.0334
	0.75	0.0000	0.0223	0.0000	0.4739	0.0000	0.1731	0.0735	0.6272	0.6601	0.4738	0.5119	0.0443
10	0.25	1.0000	0.8364	0.0000	1.0000	1.0000	1.0000	0.7840	1.0000	0.7978	0.7000	0.8765	0.8234
	0.50	1.0000	0.8092	0.0000	1.0000	1.0000	1.0000	0.8138	1.0000	0.7753	0.6503	0.8125	0.7294
	0.75	0.9419	0.7409	0.0000	0.8116	0.9456	0.5966	0.0029	0.7432	0.7464	0.6043	0.6991	0.6402

\* I = Number of indicators per factor; C = Interfactor correlation.

When each factor had 5 indicators, the accuracy rates were generally lower for all methods across interfactor correlation conditions. The only exception to this pattern was CNG at correlations of 0.25 or 0.50, for which the accuracy rate was 1.00. In addition, EKC, CD and Combined had accuracy rates of 0.99 for a correlation of 0.25. As the interfactor correlation increased in value, the accuracy rates for all of the methods studied here declined in the 5-indicator condition. At the highest interfactor correlation value, MIRT, NOHARM, and CNG had the highest accuracy rates in the 5-indicators case. The SMT had the lowest rates across all conditions in Table 1.

The proportion of replications with the correct number of factors identified by number of indicators, sample size, and method appear in Table 2. Across methods, the rate of correct factor number identification was higher for larger samples and more indicators. For 5 indicators, the highest accuracy rates were associated with CNG, except for when  $N = 2000$ . In this latter case, PA had an accuracy rate of 1.00, with CD being the second most accurate, followed by CNG. When there were 10 indicators per factor, EKC and the Combined method had the highest accuracy rates. Across conditions, SMT had the lowest rates of accurately determining the number of factors to retain. Finally, when there were 3 factors, each method studied here had higher accuracy rates for retaining the number of factors when the loadings were 0.8 as opposed to 0.6 (Table 3). This result was most marked for MAP and CD, and least so for MIRT and NOHARM.

**Table 2.** Proportion of replications with correct number of factors, number of indicators, and sample size: 3 factors.

I *	N	EKC	Hull	SMT	CD	Comb	MAP	NMREG	CNG	PA	RPA	MIRT	NOHARM
5	200	0.5073	0.3387	0.0000	0.5747	0.5073	0.5109	0.0828	0.9181	0.5000	0.4000	0.5126	0.0440
	500	0.6426	0.3965	0.0000	0.8087	0.6426	0.4387	0.3363	0.9117	0.5545	0.5004	0.5509	0.0473
	1000	0.6648	0.4471	0.0000	0.8873	0.6648	0.3657	0.4634	0.9043	0.7692	0.7122	0.5600	0.0409
	2000	0.6667	0.4541	0.0000	0.9560	0.6667	0.0000	0.5527	0.9326	1.0000	0.9124	0.5795	0.0257
10	200	0.9567	0.6678	0.0000	0.8750	0.9617	0.8683	0.4343	0.8523	0.6417	0.5834	0.7457	0.6408
	500	0.9678	0.7255	0.0000	0.8860	0.9678	0.8767	0.4476	0.8656	0.7667	0.6247	0.7845	0.6780
	1000	0.9980	0.8913	0.0000	0.9878	0.9980	0.8800	0.5923	0.8758	0.8918	0.8000	0.8156	0.7632
	2000	1.0000	0.8974	0.0000	1.0000	1.0000	0.8372	0.6601	0.9093	0.9925	0.9413	0.8481	0.8155

\* I = Number of indicators per factor; N = Sample size.

**Table 3.** Proportion of replications with correct number of factors, by number of number of factors and factor loading value.

F *	L	EKC	Hull	SMT	CD	Comb	MAP	NMREG	CNG	PA	RPA	MIRT	NOHARM
1	0.60	1.0000	0.9874	0.0000	1.0000	1.0000	1.0000	0.1394	1.0000	0.8729	0.7775	0.5933	0.2946
	0.80	1.0000	0.9943	0.0000	1.0000	1.0000	1.0000	0.0955	1.0000	0.9011	0.8383	0.6506	0.6423
3	0.60	0.7815	0.5533	0.0000	0.7878	0.7828	0.3939	0.3979	0.8561	0.7249	0.6670	0.6423	0.3473
	0.80	0.8413	0.6794	0.0000	0.9561	0.8413	0.9148	0.4861	0.9267	0.8334	0.8000	0.6915	0.3792

\* F = Number of factors; L = Factor loading value.

### 5.1.2. Mean Number of Factors Retained

The number of factors retained by method, number of indicators, and interfactor correlation in the 3-factor condition appear in Table 4. Across conditions, SMT, PA, and RPA all overestimated the number of factors to retain. In addition, NMREG also overestimated the number of factors present when the interfactor correlation was 0.5 or 0.75. Among the other methods, the mean number of factors retained was closest to the actual value of 3 for EKC, CD, Combined, CNG, MAP, MIRT, and NOHARM in the 10-indicator case, particularly with correlations below 0.75. At this highest correlation value, EKC, Combined, MIRT, and NOHARM were closest to the actual data generating factor number. With 5 indicators, the interfactor correlation had a greater impact on the performance of the methods studied here. When the correlation was 0.25 or 0.50, CNG suggested retaining 3 factors on average, whereas EKC, CD, and Combined all had a mean number of factors retained of 2.8 or higher. At the highest interfactor correlation, EKC, CD, and Combined all suggested that fewer factors be retained, with means of 2 or fewer. In contrast, CNG had a mean of 3.4 factors with 5 indicators and an interfactor correlation of 0.75, suggesting that it tended to overfactor when the latent traits were most strongly related to one another. Finally, in the 5-indicators condition, MIRT tended to overfactor

and tended to underfactor. NOHARM

**Table 4.** Number of factors retained by number of indicators, and interfactor correlation: 3 factors.

I*	C	EKC	Hull	SMT	CD	Comb	MAP	NMREG	CNG	PA	RPA	MIRT	NOHARM
5	0.25	2.9951	2.4791	7.6664	2.9942	2.9951	2.3386	2.4569	3.0000	4.0836	4.7493	3.4435	1.4236
	0.50	2.8416	2.1514	7.7285	2.8583	2.8416	1.9337	4.1305	3.0000	3.877	4.6005	3.3949	1.3478
	0.75	1.1713	1.0601	7.5671	2.0438	1.1713	1.4411	6.2610	3.4167	3.4621	4.4475	3.4835	1.3869
10	0.25	3.0000	2.7555	20.3450	3.0000	3.0000	3.0000	2.7977	3.0000	3.7758	3.6103	3.1186	3.1470
	0.50	3.0000	2.5921	20.3086	3.0000	3.0000	3.0000	5.6286	3.0000	3.4600	3.5881	3.1539	3.1386
	0.75	2.9281	2.5041	20.0646	2.6456	2.9419	2.3502	14.7273	3.2754	3.2063	3.3914	2.9699	2.8878

\* I = Number of indicators per factor; C = Interafactor correlation.

The results in Table 5 show the mean number of factors retained for each method by the number of indicators and the sample size. Many of these patterns are similar to those in Table 4 and thus will not be repeated here. Of particular import in this table is that the mean number of factors retained was closest to the population value of 3 with larger samples for EKC, CD, and Combined. In addition, the mean number of factors for PA was approximately 3 for sample size of 2000, regardless of the number of indicators. The method with the mean number of factors consistently closest to the population value of 3 across conditions was CNG. The mean number of latent traits retained by method and factor loadings for the 3-factor condition appear in Table 6. As noted above, the mean number of factors to retain was closest to the population value of 3 for CNG across factor loading values, with MIRT having the next closest mean number of factors.

**Table 5.** Number of factors retained by number of indicators, and sample size: 3 factors.

I*	N	EKC	Hull	SMT	CD	Comb	MAP	NMREG	CNG	PA	RPA	MIRT	NOHARM
5	200	2.1872	1.8383	7.9188	2.2419	2.1872	2.1154	5.1066	3.0905	4.1324	4.5296	3.4574	1.3956
	500	2.3565	1.8912	7.6964	2.6799	2.3565	2.0224	4.1902	3.0956	3.5573	3.8880	3.4409	1.4396
	1000	2.4290	1.9430	7.6782	2.8345	2.4290	1.9041	3.8438	3.1015	3.2825	3.4298	3.4242	1.3893
	2000	2.4066	1.9331	6.9912	2.9121	2.4066	1.2475	3.6984	3.0674	3.0075	3.0885	3.4400	1.3198
10	200	2.9383	2.4122	20.5598	2.7500	2.9567	2.7950	9.4493	3.1701	3.8966	4.0295	2.9323	2.8738
	500	2.9678	2.4755	20.6500	2.7836	2.9678	2.8099	9.3135	3.1651	3.3155	3.6445	3.1063	3.0665
	1000	2.9980	2.7863	20.2742	2.9939	2.9980	2.7988	7.2015	3.1261	3.2450	3.3256	3.1519	3.1230
	2000	3.0000	2.7948	19.4735	3.0000	3.0000	2.7299	4.9072	3.0944	3.0000	3.0110	3.1844	3.2781

\* I = Number of indicators per factor; N = Sample size.

**Table 6.** Number of factors retained by number of number of factors and factor loading value.

F*	L	EKC	Hull	SMT	CD	Comb	MAP	NMREG	CNG	PA	RPA	MIRT	NOHARM
1	0.60	1.0000	1.0126	7.1464	1.0000	1.0000	1.0000	5.8678	1.0000	1.8111	1.9648	1.4353	2.1216
	0.80	1.0000	1.0057	8.6957	1.0000	1.0000	1.0000	7.1758	1.0000	1.3167	1.4711	1.4743	1.4117
3	0.60	2.6332	2.1682	13.3974	2.6275	2.6378	1.9479	7.0629	3.1564	3.6379	3.8941	3.3000	1.9109
	0.80	2.7277	2.4100	15.4734	2.9228	2.7277	2.8593	5.0314	3.0824	3.3460	3.5803	3.2488	2.3896

\* F = Number of factors; L = Factor loading value.

### 5.2. 1-Factor

#### 5.2.1. Factor Retention Accuracy Rate

For the 1-factor condition, ANOVA identified the interaction of factor retention method by the number of indicators per factor and sample size ( $F_{66,286} = 2.91, \eta^2 = 0.54$ ), and method by factor loading magnitude ( $F_{11,286} = 3.01, \eta^2 = 0.17$ ) as being statistically significantly related to the factor retention accuracy rate. A number of methods exhibited accurate rates of 1.00 across sample size and number of indicators, including EKC, CD, Combined, MAP, and CNG. In contrast, SMT, NMREG, and RPA had the lowest accuracy rates across number of indicators and sample sizes. Accuracy rates were generally higher in the 10-indicator condition, particularly for MIRT and NOHARM. The accuracy rates by factor loading for the 1-factor case appears in Table 7. For the most accurate

methods, EKC, CD, MAP, CNG, and Combined, the loading values were not associated with the proportion of cases for which 1 factor was correctly identified. Among the other methods, accuracy for PA and NOHARM was most strongly impacted by loading value, where accuracy was markedly higher for the 0.80 condition.

**Table 7.** Proportion of replications with correct number of factors, number of indicators, and sample size: 1 factor.

I *	N	EKC	Hull	SMT	CD	Comb	MAP	NMREG	CNG	PA	RPA	MIRT	NOHARM
5	200	1.0000	0.9848	0.0000	1.0000	1.0000	1.0000	0.1104	1.0000	0.6032	0.5712	0.4469	0.1762
	500	1.0000	0.9872	0.0000	1.0000	1.0000	1.0000	0.1708	1.0000	0.6719	0.5004	0.4777	0.1985
	1000	1.0000	0.9931	0.0000	1.0000	1.0000	1.0000	0.2704	1.0000	0.8003	0.7122	0.4020	0.3272
	2000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	0.2802	1.0000	1.0000	0.9124	0.4886	0.4463
10	200	1.0000	0.9774	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	0.6890	0.6293	0.8507	0.6884
	500	1.0000	0.9938	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	0.8114	0.7900	0.8162	0.7029
	1000	1.0000	0.9995	0.0000	1.0000	1.0000	1.0000	0.0385	1.0000	0.9171	0.8657	0.7870	0.6860
	2000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	0.0473	1.0000	1.0000	0.9703	0.6781	0.6960

\* I = Number of indicators per factor; N = Sample size.

### 5.2.2. Mean Number of Factors Retained

The mean number of factors retained by factor loading in the 1-factor case appears in Table 8. EKC, CD, MAP, CNG, and Combined had mean number of factors retained of 1, matching the population value of 1. In addition, the Hull technique had a mean just above 1. The mean number of factors retained by method, sample size, and method in the 1-factor case appear in Table 8. In keeping with the patterns described above, EKC, CD, MAP, CNG, and Combined had a mean number of factors retained of 1 across conditions, with the mean for Hull also being very close to 1. NOHARM and MIRT both performed better (i.e., had a mean closer to 1) for more indicators, as well as for larger samples in the case of MIRT.

**Table 8.** Number of factors retained by number of indicators, and sample size: 1 factor.

I *	N	EKC	Hull	SMT	CD	Comb	MAP	NMREG	CNG	PA	RPA	MIRT	NOHARM
5	200	1.0000	1.0152	4.3520	1.0000	1.0000	1.0000	3.2768	1.0000	1.8438	2.3651	1.6519	2.2753
	500	1.0000	1.0128	4.3282	1.0000	1.0000	1.0000	2.8922	1.0000	1.4513	1.8192	1.6595	2.2958
	1000	1.0000	1.0000	4.2028	1.0000	1.0000	1.0000	2.5845	1.0000	1.1516	1.3949	1.7623	2.0383
	2000	1.0000	1.0069	4.2446	1.0000	1.0000	1.0000	2.5882	1.0000	1.0000	1.3629	1.6616	1.7961
10	200	1.0000	1.0062	12.2284	1.0000	1.0000	1.0000	11.6358	1.0000	1.4938	1.6296	1.3575	1.3419
	500	1.0000	1.0062	12.0617	1.0000	1.0000	1.0000	11.3281	1.0000	1.2728	1.5142	1.2184	1.3425
	1000	1.0000	1.0226	11.6353	1.0000	1.0000	1.0000	9.7962	1.0000	1.0996	1.2907	1.1838	1.3406
	2000	1.0000	1.0000	11.0903	1.0000	1.0000	1.0000	8.7266	1.0000	1.0000	1.0712	1.1628	1.3475

\* I = Number of indicators per factor; N = Sample size.

## 6. Discussion

The goal of this study was to investigate a wide array of statistical methods for determining the number of factors to retain when using EFA with dichotomous indicators. With respect to the statistical methods under investigation in this work, the results presented above suggest that researchers should consider using the Combined approach based on EKC, Hull, and CD. This result was in keeping with Auerswald and Moshagen [2], who also found that a combination approach was effective in ascertaining the number of factors to retain when the indicators were normally distributed. In addition, the CNG method alone also proved to be very effective in revealing the number of factors to retain. Indeed, in the most challenging case with 5 indicators per factor and an interfactor correlation of 0.75, CNG yielded the most accurate results across methods. However, when 10 indicators were present, CNG was not the most accurate method in the 0.75 interfactor correlation condition. Thus, it cannot be uniformly recommended for use in all cases.

When 10 indicators per factor were found to be present, multiple methods yielded highly accurate results for determining the number of factors to retain. Specifically, both the EKC and Combination techniques were highly accurate regardless of the interfactor correlation. In addition, MAP, CD, and CNG yielded perfectly accurate results for the 0.25 and 0.50 interfactor correlation

conditions. However, when the correlations were 0.75, they were somewhat less accurate than EKC and Combination.

When a single factor was present in the population, a number of methods studied here yielded highly accurate results, including EKC, CD, Combination, MAP, and CNG. Thus, when a single latent trait underlies the observed indicators, researchers can have confidence that any of these methods is likely to yield accurate results regarding the number of factors to retain, assuming conditions similar to those used in this study. Of course, in reality the data analyst will not know how many factors are actually present. Nonetheless, it does appear that if theory suggests a single factor is present and the methods mentioned above as being most accurate are used and yield a 1-factor solution, the researcher may be fairly confident in this result.

Several methods did not perform particularly well and may not be optimal for use in conditions similar to those used in this simulation study. In particular, PA and RPA, both of which performed well in previous studies [10], did not do so well under most conditions used in this study. Both methods tended to overfactor the data, leading to the low accuracy rates described above. It should be noted that in the previous work focusing on PA/RPA, there were more indicators per factor than were used in the current study. Thus, it seems that when the number of indicators per factor is 5, 10, or 20, neither PA nor RPA provides as accurate results as some other methods studied here, or as they do themselves with 30 or 60 indicators [10]. Similarly, the chi-square test based upon NOHARM also yielded lower accuracy rates than a number of the other techniques used in this study, particularly with 5 indicators. However, these results were due to a tendency to underfactor in the 3 latent trait condition, rather than overfactor, as was the case for PA/RPA. The MIRT model approach also yielded somewhat lower accuracy rates than a number of other methods in the 5 indicators 3-factors condition, but with a tendency to overfactor. When a single factor was present, PA, RPA, MIRT, and NOHARM all yielded overfactoring results. Taken together, these findings suggest that when statistical methods did not recommend retaining the correct number of factors, there were multiple ways in which they reached this incorrect decision. Sometimes overfactoring was the cause, whereas in other cases it was due to underfactoring. In any case, these methods may not be optimal for use with dichotomous indicators, particularly in light of the fact that other approaches (e.g., CNG, Combination and EKC, CD, or MAP in some cases) do appear to yield accurate results.

## 7. Implications for Practice

As described above, the results presented here provide some guidelines for researchers and data analysts. First, researchers and data analysts should consider the combined approach based on EKC, Hull, and CD for determining the number of factors to retain. The CNG technique would be a reasonable second choice for this purpose. Second, having more indicators per factor (presuming loadings in the range simulated here) will yield more accurate factor retention results. Thus, all other things being equal, researchers should seek to use longer scales. When scales are relatively short (e.g., 5 indicators per factor) and the interfactor correlation is high (e.g., 0.75), researchers may find it difficult to identify the correct number of factors to retain in an EFA, no matter the methodology used. Finally, larger samples were associated with more accurate results, in general. Therefore, researchers using EFA should attempt to obtain samples of 500 or more, particularly if they are working with a relatively small number (e.g., 5) of indicators per factor.

## 8. Limitations of This Study

As with all studies, there are limitations to the current work that future research should address. First, not all possible conditions that might be seen in practice were included in the current study. The goal of this work was to focus on dichotomous indicators such as might be encountered in many survey or testing environments. However, it is certainly the case that many scales rely on polytomous items that yield ordinal, rather than dichotomous, data. Thus, a limitation of this study was the fact that only items with two categories were included. Future work should include a focus on ordinal items. In addition, this study focused on cases with either 1 or 3 latent traits. However, many situations in practice involve scales with more possible latent traits, including assessments of cognitive ability, mood, or executive functioning. Thus, future work should include more latent traits while also examining a relatively small number of indicators, as was the case in this study. The current study included only latent traits that followed the standard normal distribution, which will not be present in all research contexts. For this reason, future work should consider cases where the latent trait is not normally distributed but rather might be skewed and/or highly kurtotic. Another limitation of the current study was with respect to the number of indicators per factor. In the current study, there were either 5 or 10 indicators per factor. In some instances, scales may be comprised

of more indicators than this, e.g., 20, 30, or even more. Thus, future work should examine the performance of these methods with more indicators per factor.

## 9. Conclusions

The current study was designed to extend upon prior studies comparing methods for determining the number of factors to retain in the context of EFA. This earlier work focused on a wide array of methods in the context of normally distributed indicators [2] or PA/RPA with dichotomous indicators [10]. This study included a wide array of techniques in the context of dichotomous indicators with a small number of indicators per factor. The results suggest that data analysts and researchers might be best served to use CNG when there are 5 indicators per factor, and CNG, EKC, MAP, or the combination techniques for 10 indicators per factor. In such cases, one can anticipate accurate results regarding the number of factors to retain, particularly when the interfactor correlations are 0.5 or lower. Of course, regardless of what the statistical methods might suggest, researchers should always ensure that their ultimate findings are grounded in theory. If the optimal factor retention solution based on statistical results does not yield a result that is theoretically sound, the researcher should reconsider their analyses. However, when used in conjunction with theory, several methods studied here can help researchers identify the optimal number of factors to retain from an EFA.

**Funding:** This research received no external funding.

**Data Availability Statement:** Simulation code available upon request.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

- Gorsuch, R.L. *Factor Analysis*, 2nd ed.; Lawrence Erlbaum Associates, Publishers: Hillsdale, NJ, USA, 1983.
- Auerswald, M.; Moshagen, M. How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychol. Methods* **2019**, *24*, 468–491. [[CrossRef](#)] [[PubMed](#)]
- Boltzmann, M.; Schmidt, S.B.; Gutenbrunner, C.; Krauss, J.K.; Hoglinger, G.U.; Weimar, C.; Rollnik, J.D. Validity of the Early Functional Ability scale (EFA) among critically ill patients undergoing early neurological rehabilitation. *BMC Neurol.* **2022**, *22*, 333. [[CrossRef](#)]
- Selau, T.; da Silva, M.A.; de Mendonca Filho, E.J.; Bandeira, D.R. Evidence of validity and reliability of the adaptive functioning scale for intellectual disability (EFA-DI). *Psicol. Reflex. E Crit.* **2020**, *33*, 26. [[CrossRef](#)] [[PubMed](#)]
- Beck, A.T.; Steer, R.A.; Garbin, M.G. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin. Psychol. Rev.* **1988**, *8*, 77–100. [[CrossRef](#)]
- Beck, A.T.; Epstein, N.; Brown, G.; Steer, R.A. An inventory for measuring clinical anxiety: Psychometric properties. *J. Consult. Clin. Psychol.* **1988**, *56*, 893–897. [[CrossRef](#)]
- Selbom, M. The MMPI-2-Restructured Form (MMPI-2-RF): Assessment of personality and psychopathology in the Twenty-First Century. *Annu. Rev. Clin. Psychol.* **2019**, *15*, 149–177. [[CrossRef](#)]
- Kuncel, N.R.; Wee, S.; Serafin, L.; Hezlett, S.A. The validity of the Graduate Record Examination for Master's and Doctoral programs: A meta-analytic investigation. *Educ. Psychol. Meas.* **2010**, *70*, 340–352. [[CrossRef](#)]
- Wechsler, D. *Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV)*; American Psychological Association: Washington, DC, USA, 2008.
- Guo, W.; Choi, Y.-J. Assessing dimensionality of IRT models using traditional revised parallel analyses. *Educ. Psychol. Meas.* **2023**, *83*, 609–629. [[CrossRef](#)]
- Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*; Pearson: New York, NY, USA, 2019.
- Finch, W.H. *Exploratory Factor Analysis*; Sage: Thousand Oaks, CA, USA, 2019.
- Bollen, K.A. *Structural Equations with Latent Variables*; John Wiley & Sons: New York, NY, USA, 1989.
- Tong, X.; Bentler, P.M. Evaluation of a New Mean Scaled and Moment Adjusted Test Statistic for SEM. *Struct. Equ. Model.* **2013**, *20*, 148–156. [[CrossRef](#)]
- Kim, J.-O.; Mueller, C.W. *Factor Analysis: Statistical Methods and Practical Issues*; Sage: Thousand Oaks, CA, USA, 1978.
- Brown, T.A. *Confirmatory Factor Analysis for Applied Research*; The Guilford Press: New York, NY, USA, 2015.
- Hayashi, K.; Bentler, P.M.; Yuan, K.-H. On the likelihood ratio test for the number of factors in exploratory factor analysis. *Struct. Equ. Model.* **2007**, *14*, 505–526. [[CrossRef](#)]
- Horn, J.L. A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika* **1965**, *30*, 179–185. [[CrossRef](#)]
- Agresti, A. *Categorical Data Analysis*; Jon Wiley & Sons: New York, NY, USA, 2013.
- Green, S.B.; Levy, R.; Thompson, M.S.; Lu, M.; Lo, W.-J. A Proposed Solution to the Problem with using Completely Random Data to Assess the Number of Factors with Parallel Analysis. *Educ. Psychol. Meas.* **2012**, *72*, 357–374. [[CrossRef](#)]
- Ruscio, J.; Roche, B. Determining the Number of Factors to Retain in an Exploratory Factor Analysis using Comparison Data of Known Factorial Structure. *Psychol. Assess.* **2012**, *24*, 282–292. [[CrossRef](#)]

22. Velicer, W.F. Determining the Number of Components from the Matrix of Partial Correlations. *Psychometrika* **1976**, *41*, 321–327. [[CrossRef](#)]
23. Caron, P.-O. Minimum Average Partial Correlation and Parallel Analysis: The Influence of Oblique Structures. *Commun. Stat.-Simul. Comput.* **2019**, *40*, 2110–2117. [[CrossRef](#)]
24. Garrido, L.E.; Abad, F.J.; Ponsoda, V. Performance of Velicer’s Minimum Average Partial Factor Retention Method with Categorical Variables. *Educ. Psychol. Meas.* **2011**, *71*, 551–570. [[CrossRef](#)]
25. Zwick, W.R.; Velicer, W.F. Comparison of Five Rules for Determining the Number of Components to Retain. *Psychol. Bull.* **1986**, *99*, 432–442. [[CrossRef](#)]
26. Cattell, R.B. The Scree Test for the Number of Factors. *Multivar. Behav. Res.* **1966**, *2*, 245–276. [[CrossRef](#)]
27. Raiche, G.; Walls, T.A.; Magis, D.; Riopel, M.; Blais, J.-G. Non-Graphical Solutions for Cattell’s Scree Test. *Methodology* **2012**, *9*, 23–29. [[CrossRef](#)]
28. Zoski, K.W.; Jurs, S. Using Multiple Regression to Determine the Number of Factors to Retain in Factor Analysis. *Mult. Linear Regres. Viewp.* **1993**, *20*, 5–9.
29. Lorenzo-Seva, U.; Timmerman, M.E.; Kiers, H.A. The Hull method for selecting the number of common factors. *Multivar. Behav. Res.* **2011**, *46*, 340–364. [[CrossRef](#)]
30. Braeken, J.; van Assen, M.A. An empirical Kaiser criterion. *Psychol. Methods* **2017**, *22*, 450–466. [[CrossRef](#)] [[PubMed](#)]
31. Marcenko, V.A.; Pastur, L.A. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Shornik* **1967**, *1*, 457–483. [[CrossRef](#)]
32. McDonald, R.P. *Nonlinear Factor Analysis*; Psychometric Monographs, No. 15; Psychometric Society: Richmond, VA, USA, 1967.
33. Gessaroli, M.E.; De Champlain, A.F. Using an approximate Chi-square statistic to test the number of dimensions underlying the responses to a set of items. *J. Educ. Meas.* **1996**, *33*, 157–179. [[CrossRef](#)]
34. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychom.* **1981**, *46*, 443–459. [[CrossRef](#)]
35. Fabrigar, L.R.; Wegener, D.T. *Exploratory Factor Analysis*; Oxford University Press: Oxford, UK, 2011.
36. Green, S.; Xu, Y.; Thompson, M.S. Relative accuracy of two modified parallel analysis methods that use the proper reference distribution. *Educ. Psychol. Meas.* **2018**, *78*, 589–604. [[CrossRef](#)]
37. Green, S.B.; Thompson, M.S.; Levy, R.; Lo, W.-J. Type I and II Error Rates and Overall Accuracy of the Revised Parallel Analysis Method for Determining the Number of Factors. *Educ. Psychol. Meas.* **2015**, *75*, 428–457. [[CrossRef](#)]
38. Lim, S.; Jahng, S. Determining the number of factors using parallel analysis and its recent variants. *Psychol. Methods* **2019**, *24*, 452–467. [[CrossRef](#)]
39. Preacher, K.J.; MacCallum, R.C. Repairing Tom Swift’s Electric Factor Analysis Machine. *Underst. Stat.* **2003**, *2*, 13–43. [[CrossRef](#)]
40. Green, S.B.; Redell, N.; Thompson, M.S.; Levy, R. Accuracy of Revised and Traditional Parallel Analyses for Assessing Dimensionality with Binary Data. *Educ. Psychol. Meas.* **2016**, *76*, 5–21. [[CrossRef](#)]
41. Finch, W.H.; Habing, B. Performance of DIMTEST and NOHARM based statistics for testing unidimensionality. *Appl. Psychol. Meas.* **2007**, *31*, 292–307. [[CrossRef](#)]
42. Finch, W.H.; Habing, B. Comparison of NOHARM and DETECT: Counting dimensions and allocating items. *J. Educ. Meas.* **2005**, *42*, 149–170. [[CrossRef](#)]
43. Svetina, D.; Levy, R. Dimensionality in compensatory MIRT when complex structure exists: Evaluation of DETECT and NOHARM. *J. Exp. Educ.* **2016**, *84*, 398–420. [[CrossRef](#)]
44. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
45. Orcan, F. MonteCarloSEM: An R package to simulate data for SEM. *Int. J. Assess. Tools Educ.* **2021**, *8*, 704–713. [[CrossRef](#)]
46. Worthington, R.L.; Whittaker, T.A. Scale development research: A content analysis and recommendations for best practices. *Couns. Psychol.* **2006**, *34*, 806–838. [[CrossRef](#)]
47. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates, Publishers: Hillsdale, NJ, USA, 1988.
48. Raiche, G.; Magis, D. *nFactors: Parallel Analysis and Other Non-Graphical Solutions to the Cattell Scree Test*. An R Software Library, R package version 2.4.1.1. 2022.
49. O’Connor, B.P. *EFA.dimensions: Exploratory Factor Analysis Functions for Assessing Dimensionality*. A R Software Library, R package version 0.1.7.7. 2023.
50. Chalmers, R.P. MIRT: A multidimensional item response theory package for the R environment. *J. Stat. Softw.* **2012**, *48*, 1–29. [[CrossRef](#)]
51. Robitzsch, A. *SIRT: Supplementary Item Response Theory Models*. An R Software Library, R package version 3.13-228. 2022.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.