Psychological Test and Assessment Modeling, Volume 56, 2014 (1), 104-123

Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis

Russell T. Warne[1] *& Ross Larsen*[2]

Abstract

Exploratory factor analysis (EFA) is a widely used statistical method in which researchers attempt to ascertain the number and nature of latent factors that explain their observed variables. When conducting an EFA, researchers must choose the number of factors to retain – a critical decision that has drastic effects if made incorrectly. In this article, we examine a newly proposed method of choosing the number of factors to retain. In the new method, confidence intervals are created around each eigenvalue and factors are retained if the entire eigenvalue is greater than 1.0. Results show that this new method outperforms the traditional Guttman rule, but does not surpass the accuracy of Velicer's minimum average partial (MAP) or Horn's parallel analysis (PA). MAP was the most accurate method overall, although it had a tendency to underfactor in some conditions. PA was the second most accurate method, although it frequently overfactored. PA was also found to be sensitive to sample size and MAP was found to occasionally grossly overfactor; these findings had not previously been reported in the literature.

Key words: exploratory factor analysis, simulation study, Monte Carlo study, principal components analysis

[1] *Correspondence concerning this article should be addressed to:* Dr. Russell T. Warne, Department of Behavioral Science, Utah Valley University, 800 W. University Parkway, Mail Code 115, Orem, UT 84058, USA; email: rwarne@uvu.edu

[2] University of Virginia

Methodologists largely agree that among the choices that a researcher makes in conducting an exploratory factor analysis (EFA) or principal components analysis (PCA), the decision of the number of factors to retain is among the most important (Costello & Osborne, 2005; Fava & Velicer, 1996; Thompson, 2004; Thompson & Daniel, 1996; Zwick & Velicer, 1986). Indeed, O'Connor stated, ". . . the crucial decision is that of determining how many factors to retain. Assorted decisions on the other issues generally produce similar results when the optimal number of factors is specified" (2000, p. 396; see also Shönemann, 1990).

The importance of retaining the correct number of factors cannot be overstated. EFA and PCA, for example, are often used in the construction and evaluation of psychometric instruments (e.g., Dombrowski, Watkins, & Brogan, 2009; Falk, Lind, Miller, Piechowski, & Silverman, 1999; Zheng, Hall, Dugan, Kidd, & Levine, 2002). If a researcher retains the improper number of factors when constructing an instrument, then the interpretation of a psychometric test could become severely distorted. Under these circumstances, construct validity would likely be compromised, and researchers would unwittingly have an incorrect understanding of the factor structure of the data produced by items on a psychometric instrument. This may lead researchers to have an oversimplified understanding of the construct (in the case of underfactoring) or unnecessarily complex results, which would violate the scientific principle of parsimony (in the case of overfactoring).

EFA and PCA are also used as a method to convert a large number of observed variables into a smaller number of artificial variables to make later data analysis more manageable (Thompson, 2004). Again, if a researcher retains the improper number of factors when reducing data, the later analyses could produce inaccurate results and thus lead researchers astray in their conclusions.

When discussing the issue of selecting the number of articles in a previous article, we explained to readers how they could estimate confidence intervals (CIs) for eigenvalues in an exploratory factor analysis (Larsen & Warne, 2010). Using publicly available data, we showed readers how they could use two methods to calculate CIs: a formula based on the Central Limit Theorem (CLT) and bootstrapping. Both methods produce similar results, although the CLT method produced wider CIs in our example.

The equation for calculating eigenvalue CIs is relatively simple:

$$l_i \pm z_{(1-\alpha/2)} * \left(\sqrt{\frac{2l_i^2}{n}} \right)$$

where l_i represents the observed eigenvalue, $z_{(1-\alpha/2)}$ represents the appropriate z-value for the CI width (e.g., 1.96 for a 95% CI), and n represents the sample size. To help researchers who wish to calculate CIs, we provided SPSS and SAS syntax to conduct the calculations. Similarly, we provided SPSS and SAS syntax to use bootstrapping to estimate eigenvalue CIs. The SPSS syntax was based on Zientek and Thompson's (2007) work and the SAS syntax was based on Tonsor's (2004) work. Details about the syntax code for both programs are available in our previous article (Larsen & Warne, 2010).

In the discussion section of the article, we mentioned that eigenvalue CIs could be used in helping a researcher decide how many factors to retain in an exploratory factor analysis. Specifically, we stated that the Guttman rule could be modified so that retained factors are those for which the entire eigenvalue CI is greater than 1.0. Moreover, we made the claim that using eigenvalue CIs would improve the quality of factor retention decisions. The main purpose of this article is to test this claim by using simulated data in a Monte Carlo study to compare the accuracy of various, more traditional factor retention decisions with our proposed modification of the Guttman (1954) rule, in which all factors with the entire 95% CI above 1.0 are retained. The specific research questions of this study are as follows:

1. Which of the five methods (Guttman, CLT modified Guttman, bootstrap modified Guttman, parallel analysis, minimum average partial) most frequently determines the correct number of factors and produces the smallest bias in an EFA?
2. Are the modifications to the Guttman rule suggested by Larsen and Warne (2010) more accurate than the Guttman rule?
3. Which data characteristics (i.e., sample size, number of observed variables, number of factors, correlation among factors, and strength of factor loadings) have the largest impact on bias for the methods of determining the number of factors?

Notes on terminology

Before continuing, it is important to make some of our usage and terminology clear. First, we use the term "exploratory factor analysis" or "EFA" rather freely. When using the term, we also refer to PCA. We are aware that there is some controversy surrounding whether PCA is a "real" factor analysis (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Gorsuch, 2003; Velicer, Eaton, & Fava, 2000), but we do not take a position in the argument. However, we do want to acknowledge that those who disagree with our use of the phrase "exploratory factor analysis" to also refer to PCA have some strong arguments in their favor. Similarly, our use of the word "factors" could also be interpreted as referring to "components." Readers should also be aware that nothing that we discuss in this article applies to confirmatory factor analysis (CFA) because in that method the number of factors is hypothesized *a priori* based on theory or previous data (Thompson, 2004).

Second, we use the term "Guttman rule" to refer to the strategy of retaining all factors with eigenvalues greater than 1.0. This rule is also called the K1 rule (after a 1960 article by Kaiser) and the eigenvalues-greater-than-one rule. However, we prefer the term "Guttman rule" because Guttman (1954) originally proposed the method. We use the phrase "modified Guttman rule" or just "modified Guttman" to refer to our proposed modifications where factors with the entire eigenvalue CI above 1.0 are retained (Larsen & Warne, 2010). In this article, we also use the terms "CLT modified Guttman rule" and "Bootstrap modified Guttman rule" to distinguish the methods by which the CIs were calculated in a particular instance.

Methods for deciding the number of factors to retain

Although there are many methods for deciding on the number of factors to retain, this paper will focus on three methods, in addition to the two modified Guttman rules that we examine in this article. The three methods are the original Guttman rule, parallel analysis (PA), and minimum average partial (MAP).

Guttman rule

Guttman (1954) proposed that in an EFA, all factors with eigenvalues greater than 1.0 should be retained. His logic in proposing this rule was that EFA is a data reduction method, and it is reasonable to require any retained factors to explain more variance than is explained by a single variable – which will be 1.0 when all variables are standardized. The Guttman rule is probably the most popular method for determining the number of factors to retain in an EFA among psychological researchers, likely because the Guttman rule is the default method of factor retention in programs like SPSS (Fabrigar et al., 1999; Ford, MacCallum, & Tait, 1986; Hayton, Allen, & Scarpello, 2004; Henson & Roberts, 2006; Thompson & Daniel, 1996; Warne, Lazo, Ramos, & Ritter, 2012).

Since its proposal over half a century ago, the Guttman rule has been heavily criticized. Thompson (2004), for example, disapproves of the Guttman rule because the 1.0 threshold is arbitrary. Cliff (1988) convincingly showed that the Guttman rule's logic that 1.0 should be the lower bound of any retained factor only applies to population data – which researchers rarely have – and not sample data. Simulation studies also have shown that strict application of the Guttman rule tends to lead researchers to retain too many factors, especially when there are a large number of variables and the sample size is large (Hakstian, Rogers, & Catell, 1982; Velicer et al., 2000; Zwick & Velicer, 1982, 1986).

Parallel analysis

A more empirically sound method of retaining the number of factors is PA. Originally developed by Horn (1965), PA requires researchers to generate a number of datasets of random numbers that have the same distribution properties as the actual raw data and to calculate the average eigenvalues from the random data for each factor. The eigenvalues for the real data are compared to the corresponding random data eigenvalue; all factors that have the observed eigenvalue higher than the random data's eigenvalue are retained. It makes intuitive sense to use PA, because one could make a convincing argument that factors whose eigenvalues are lower than an eigenvalue from random data would be meaningless and likely due to mere chance correlations among variables.

In the decades since Horn (1965) proposed PA, support has grown for the practice. Simulation studies showed that PA performed well under reasonable conditions, although – like the Guttman rule – it tended to retain too many factors (Zwick & Velicer, 1986). In response, modern proponents of PA suggest that researchers compare each observed eigenvalue to the 95^{th} percentile of the corresponding eigenvalue for the random datasets, which is a more conservative approach (Cota, Longman, Holden, Fekken, & Xinaris, 1993; Glorfeld, 1995; O'Connor, 2000). Although PA is still not available in common statistical analysis software, O'Connor (2000) has published SPSS and SAS syntax that makes PA easy to perform. Liu and Rijmen (2008) have produced SAS syntax that permits PA to be conducted with ordinal data. For this simulation study we used O'Connor's (2000) SAS syntax to make factor retention decisions using PA.

Minimum average partial

Velicer (1976) proposed the MAP method of deciding the number of factors to retain in an EFA. Unlike PA or the Guttman rule, MAP does not use observed eigenvalues to help researchers make the retention decision. Instead, MAP determines the number of factors by comparing the systematic and unsystematic variance remaining in a correlation matrix after each additional factor is extracted. This process continues until the average squared partial correlation is at a minimum.

MAP is not commonly used, despite performing well in simulation studies (e.g., Velicer et al., 2000; Zwick & Velicer, 1982, 1986). As with PA, O'Connor (2000) has produced SPSS and SAS syntax to help researchers apply MAP to their EFAs; the SAS syntax from O'Connor (2000) was used for this simulation study.

Methods

Simulated conditions

Five major variables were manipulated in this Monte Carlo study: (a) the number of observed variables in the dataset, (b) the number of factors present in the population matrix, (c) the sample size, (d) the correlation among population factors, and (e) the strength of the factor loadings. The correlation matrix was calculated by using classic structural equation modeling equations, specifically:

$$x = \Lambda_x \xi + \delta$$

where x is the observed (endogenous) variable, Λ_x are the factor loadings (or pattern matrix), ξ is the true score associated with the factor, and δ is the error component associated with the exogenous variable. Thus the variance/covariance matrix can be defined as:

$$\Sigma_{xx} = \Lambda_x \Phi \Lambda_x^{'} + \varepsilon$$

where again, Λ_x is the factor loading and Φ is the variance/covariance matrix of the factors, and ε is the error term. Multiplying Λ_x by Φ will result in the vector of structure coefficients (For further details on the theory of structural equation modeling, we invite readers to consult Kline, 2005). Sampling variability was introduced into the simulated data by using the RANDNORMAL function in SAS (SAS, 2009), which produces random samples drawn from a multivariate normally distributed population. The eigenvalues were calculated from the simulated data using principal components analysis using the SAS PROC PRINCOMP (SAS, 2009).

We decided to use structural equation modeling instead of multidimensional item response theory (IRT) to simulate our data because both EFA and PCA are special cases of SEM, and we thought that keeping the data generation and analysis within the structural equation modeling framework would simplify interpretation and remain faithful to the way that most psychometricians and quantitative psychologists think of EFA when using the method. Readers who prefer an IRT basis for simulated data for examinee responses should remember that multidimensional IRT is a special case of structural equation modeling (Reckase, 1997) and that statistics and parameter estimates in IRT can be converted into factor analysis path estimates (Raju, Laffitte, & Byrne, 2002).

Number of observed variables

For this simulation, we decided that there would be three conditions for the number of observed variables. Each condition would have 15, 30, and 45 observed variables, respectively. We chose these values because similar values had been used in previous simulation studies on EFA and PCA (e.g., Fava & Velicer, 1992, 1996). Moreover, Henson and Roberts (2006) found that of the articles using EFA that they examined, a mean of 23.73 of observed variables were factor analyzed (SD = 16.70), indicating that results of a simulation of 15-45 observed variables would generalize to most published EFAs.

Number of factors

There were also three conditions for the number of factors in our Monte Carlo simulation. We decided that the population matrices would consist of 1, 3, or 5 factors. The number of observed variables per factor was always equal within each condition, which is consistent with previous simulation studies (e.g., Fava & Velicer, 1992, 1996) and Henson and Roberts's (2006) review of published EFAs.

Sample size

We decided to define sample sizes as a ratio between research participants and observed variables. There were seven conditions for this variable: 2:1, 5:1, 8:1, 11:1, 14:1, 17:1, and 20:1. Therefore actual sample sizes ranged from 30 to 900. We wanted this diverse range of sample size so that we could observe the various factor retention decision methods under a wide variety of realistic data conditions (Henson & Roberts, 2006).

Factor correlations

We decided to simulate the degree of intercorrelation among factors with three conditions: perfectly orthogonal factors (i.e., $r = 0$), factors that are modestly correlated ($r = .25$), and distinctly correlated factors ($r = .50$). We believed that the correlation among factors was an important condition to simulate for two reasons. First, many factors derived from psychological data are at least modestly correlated (Thompson, 2004). Second, correlated factors are likely to merge together when a researcher extracts fewer factors than he or she should (Fava & Velicer, 1992). All factors within a population condition had the same intercorrelation. When only one factor is present in the population matrix, there cannot be any correlation with other factors, so this variable was not simulated in the one-factor case.

Factor loading strength

The magnitude of the factor loadings were varied, ranging from relatively weak (.3), to moderate (.5), to relatively strong (.7). We believe that these levels of factor loading strength represent an adequate range of the ratio between signal and noise in latent variables that is often found in simulated (e.g., Fava & Velicer, 1992) and empirical research.

Simulation conditions and statistical analysis

One thousand replications were run for all 441 combinations of conditions of the five independent variables. For each replication, all five methods of choosing the number of factors (i.e., the original Guttman rule, the CLT modified Guttman rule, the bootstrap Guttman rule using 100 iterations, PA, and MAP) were all used to produce an estimate of the number of factors ($\hat{\theta}$).

The dependent variable in this study is bias, which we defined as the difference between the suggested number of factors ($\hat{\theta}$) and the actual number of factors (θ). This value was calculated for each method in each condition. The results were compared across the various conditions: number of observed variables, ratio of subjects to observed variables, number of factors, factor intercorrelation, and factor loading strength. The resulting replications were analyzed with basic descriptive statistics to examine the amount and consistency of bias in each method. Regression analysis was then used to examine the relative impact that the sample conditions had on each decision method.

Results

Descriptive statistics

Table 1 shows the descriptive statistics for the bias of the five methods averaged across all levels of observed variables, sample size, number of factors, and correlations of the factors. As can be seen, the MAP technique generally outperforms the other techniques across all conditions in the study with a mean bias of -.47 and a median bias of 0. The next least biased method of determining the number of factors is PA, which had in this study a mean bias of +3.22 and a median bias of +2. The bootstrap and CLT methods performed almost equally well in the study, with means of +4.35 and +4.06 respectively; both of these modified Guttman methods had a median bias of +4. The worst method of the five we examined was the original Guttman rule, which had a mean bias of +5.47 and a median bias of +5.

The mode bias for all methods was 0, indicating that every method produced the correct number of factors more often than any other number. However, the percentage of samples that had the correct number of factors identified varied widely. The Guttman method only identified the correct number of factors in 13.5% of replications; the bootstrap method was slightly more accurate, with the correct number of factors being identified 16.1% of the time. The CLT method only identified the correct number of factors in 18.7% of samples. PA was accurate in 26.2% of replications. The only method that was correct for approximately half of all conditions was MAP, which correctly identified the number of factors 48.5% of the time.

Despite the low average bias among results using the MAP technique, the method nonetheless occasionally produced the highest levels of bias that we encountered in the simulation. As can be seen in Table 1, MAP found as many as 43 factors more than actually existed in the data. Indeed, a small number of replications (0.00067%) using MAP had a greater bias than the maximum amount of bias for *any* of the other methods. All of these cases occurred in samples that contained 30 or 45 observed variables, with the largest number of variables accounting for almost ¾ of these cases. No other sample characteristic demonstrated a clear relationship with the presence of this drastic overfactoring in MAP. Perhaps the overfactoring can be attributed to the lack of a local minimum in the MAP criterion in these samples (see Velicer et al. 2000, p. 54). However, we could find no pattern to the conditions in which MAP greatly overfactored. Thus, results that indicate an unexpectedly large number of factors from the MAP technique should be interpreted with caution.

Table 1:
Descriptive Statistics of Bias of the Five Methods

Statistic	Guttman Bias	CLT Bias	Bootstrap Bias	MAP Bias	PA Bias
Mean	5.47	4.06	4.35	-0.47	3.22
Median	5.00	4.00	4.00	0.00	2.00
Std. Deviation	4.29	3.91	3.88	2.10	3.61
Minimum	-3.00	-4.00	-4.00	-5.00	-4.00
Maximum	17.00	15.00	16.00	43.00	15.00
Mode	0.00	0.00	0.00	0.00	0.00
% with 0 Bias	13.5%	18.7%	16.1%	48.5%	26.2%

Note. 95% confidence intervals were used in the two modified Guttman rules.

Table 2:
Regression Effect Sizes and Selected Interaction

	Partial η^2				
Variable	**Guttman**	**CLT**	**Bootstrap**	**MAP**	**PA**
Factor Loading Strength	.68	.60	.63	.09	.57
Correlation	.00	.01	.01	.01	.01
Number of Factors	.50	.43	.51	.33	.18
Sample Size	.01	.04	.06	.00	.12
Number of Observed Variables	.72	.66	.68	.04	.30
Correlation*Number of Factors	.00	.01	.01	.01	.01

Note. 95% confidence intervals were used in the two modified Guttman rules.

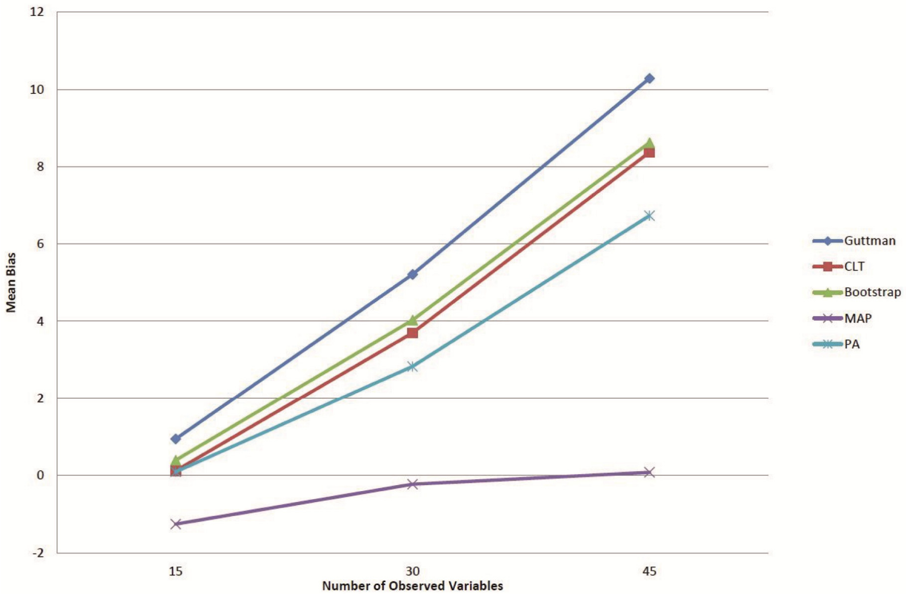

Figure 1:
Mean bias across number of observed variables for the five methods

Regression analysis

Table 2 shows the selected results from the multiple regression analyses. Although all possible two-way interactions were examined in the analysis, only the statistically significant interaction between factor intercorrelation and number of factors is shown. All other interactions are omitted for parsimony. Unsurprisingly, the partial η^2 values for the Guttman rule and its two variants are very similar – indicating that they are susceptible to the same sources of bias – factor loading strength, number of factors, and number of items. The other eigenvalue-based factor retention method, PA, is also highly influenced by the factor loading strength and the number of items, but is relatively robust to changes in the number of factors. However, PA has the greatest sensitivity to changes in sample size (partial η^2 = .12), which is important to consider because EFA is a large *n* method. MAP is the most robust of the five methods considered in this article, with only the number of factors having a large impact on bias (partial η^2 = .33). The remaining potential sources of bias all have a very small impact on bias (partial η^2 < .10). This includes the only statistically significant interaction, which was the interaction between factor intercorrelation and the number of factors. All partial η^2 values for this interaction were ≤ .01. This indicates that the influences of the independent variables are relatively independent of one another in our simulation study.

Number of observed variables

As stated in the methods section, the number of observed variables varied was 15, 30, and 45. The average biases across these settings are shown in Table 3 and Figure 2. The Guttman, CLT, Bootstrap, and PA methods all show a tendency to overfactor the data as the number of observed variables increases. The Guttman rule is the worst offender in this regard, followed by the two modified Guttman methods – which had similar results – and, finally, the PA method. The MAP method almost had the opposite pattern (with a slight tendency to underfactor when the number of observed variables was small) and approaches an unbiased condition as the number of observed variables increases.

Ratio of sample size to variables

The mean bias of the techniques as averaged across the ratio of subjects to observed variables is shown in Table 4. As can be seen in the table, there is a positive relationship between the sample size-to-variables ratio and the mean bias for all the methods, except MAP (for which the trend is reversed). This relationship also appears to be asymptotic. In other words, as the ratio of subjects to observed variables increases, the change in bias for each method becomes smaller and smaller, seeming to settle on an end amount of bias.

Table 3:
Mean Bias across Number of Observed Variables for the Five Methods

Number of Observed Variables	Guttman (SD)	CLT (SD)	Bootstrap (SD)	MAP (SD)	PA (SD)
15	0.94 (1.21)	0.12 (1.15)	0.39 (1.17)	-1.26 (2.15)	-0.09 (1.10)
30	5.21 (2.01)	3.70 (1.99)	4.03 (1.98)	-0.23 (1.70)	2.83 (2.23)
45	10.28 (2.37)	8.37 (2.51)	8.62 (2.42)	0.08 (2.16)	6.73 (3.28)

Note. 95% confidence intervals were used in the two modified Guttman rules.

Table: 4:
Mean Bias of the Ratio of Sample Size to Number of Observed Variables by Technique

Ratio	Guttman (SD)	CLT (SD)	Bootstrap (SD)	MAP (SD)	PA (SD)
2:1	4.20 (3.99)	2.61 (3.31)	3.32 (3.10)	-0.02 (2.37)	1.18 (2.37)
5:1	5.42 (4.21)	3.68 (3.67)	4.15 (3.69)	-0.37 (2.13)	2.61 (3.16)
8:1	5.48 (4.28)	4.09 (3.84)	4.40 (3.87)	-0.50 (2.06)	3.22 (3.47)
11:1	5.52 (4.33)	4.33 (3.94)	4.53 (3.97)	-0.55 (2.06)	3.57 (3.64)
14:1	5.55 (4.35)	4.48 (4.00)	4.62 (4.03)	-0.59 (2.01)	3.83 (3.77)
17:1	5.56 (4.37)	4.59 (4.05)	4.68 (4.09)	-0.62 (1.98)	3.99 (3.84)
20:1	5.57 (4.38)	4.68 (4.09)	4.74 (4.12)	-0.63 (2.02)	4.13 (3.90)

Note. 95% confidence intervals were used in the two modified Guttman rules.

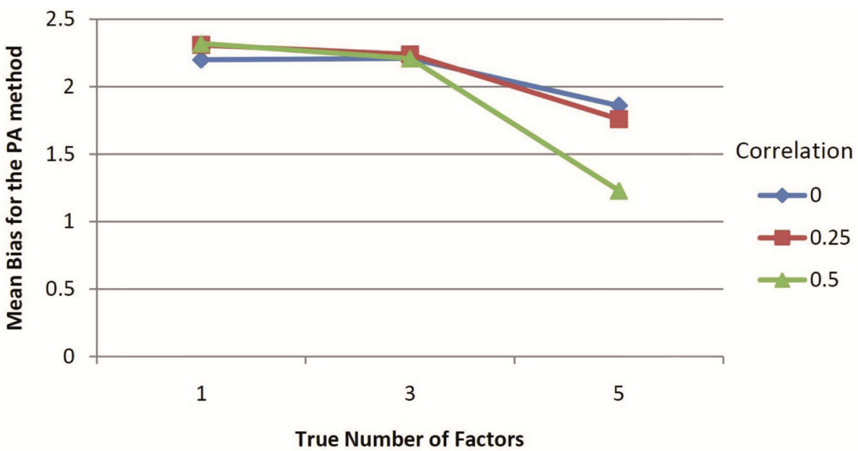

Figure 2:
Interaction between number of factors and amount of correlation between factors for the PA method

Number of factors

Table 5 reveals that all the methods showed a tendency to have lower bias as the number of factors increased. For the MAP technique, this manifests itself in a tendency to under-factor as the number of population factors increases as displayed in Figure 3. As shown previously, the number of observed variables has a tendency to overinflate the number of factors estimated for the majority of the methods. Therefore, as the number of population factors increases, the eigenvalue-based estimation techniques are more likely to retain the proper number of factors, assuming that the number of observed variables remains constant.

Factor correlation

The correlation among the factors had very little impact on the amount of bias generated by each of the methods, as seen in Table 6. There was a slight interaction between number of factors and correlation for the CLT, Bootstrap, and PA methods as illustrated by Table 2 and Figures 3 and 4. As the number of factors and correlation among factors increased, there was a tendency in these methods to have lower bias. This same pattern was also true for the MAP method, but the tendency was to underfactor as the number of factors and the correlation increased.

Table 5:
Mean Bias across the True Number of Factors

Number of Factors	Guttman (SD)	CLT (SD)	Bootstrap (SD)	MAP (SD)	PA (SD)
1	6.61 (4.41)	5.10 (4.05)	5.55 (4.04)	0.54 (1.31)	3.88 (1.84)
3	5.52 (4.17)	4.16 (3.75)	4.34 (3.74)	0.06 (1.44)	3.37 (3.50)
5	4.29 (3.91)	2.99 (3.60)	3.15 (3.46)	-2.01 (2.40)	2.40 (3.33)

Note. 95% confidence intervals were used in the two modified Guttman rules.

Table 6:
Mean Bias across the Intercorrelations among Factors

Inter-correlation	Guttman (SD)	CLT (SD)	Bootstrap (SD)	MAP (SD)	PA (SD)
0.00	5.51 (4.26)	4.13 (3.85)	4.42 (3.84)	-0.39 (2.18)	3.31 (3.55)
0.25	5.41 (4.26)	4.09 (3.89)	4.38 (3.86)	-0.35 (2.03)	3.26 (3.59)
0.50	5.42 (4.30)	3.97 (3.98)	4.24 (3.94)	-0.68 (2.06)	3.08 (3.70)

Note. 95% confidence intervals were used in the two modified Guttman rules.

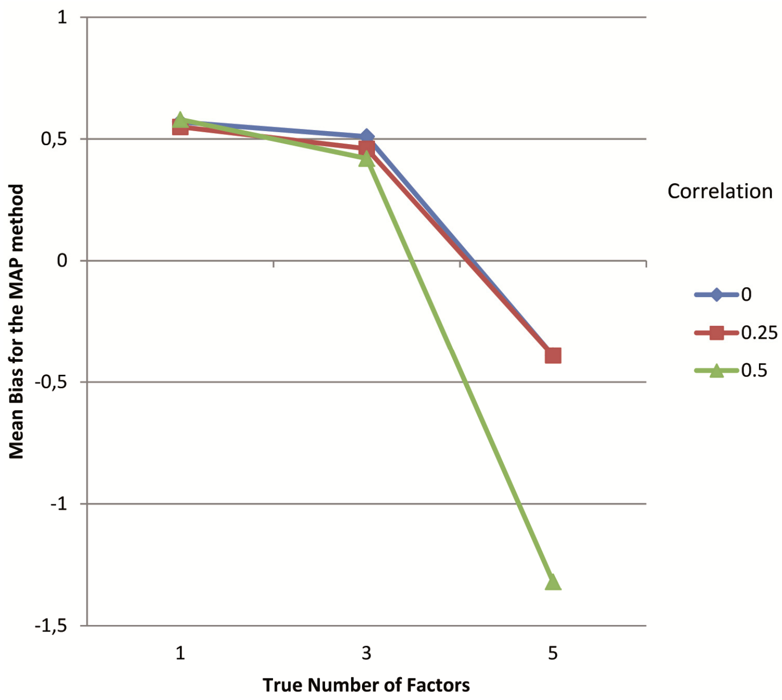

Figure 3:
Interaction between number of factors and amount of correlation between factors for the MAP method

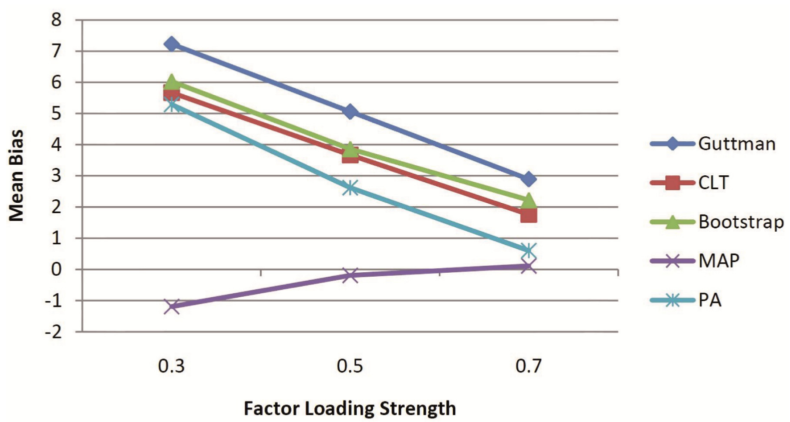

Figure 4:
Line plot of mean bias averaged across levels of factor loading strength

Factor loading strength

Table 7 and Figure 5 show the results from varying the factor loading strength at three levels: .3, .5, and .7. The results show that all methods converged closer to an average bias of 0 as the factor loading strength increases. Interestingly, the figure shows that MAP underfactors when the factor loadings are relatively weak (mean bias = -1.19), while the eigenvalue-based methods all overfactor. PA is the only eigenvalue-based method that exhibited low bias when factor loadings were .7 (mean bias = .060).

Table 7:
Mean Bias across Factor Loading Strength

Loading Strength	Guttman (SD)	CLT (SD)	Bootstrap (SD)	MAP (SD)	PA (SD)
.3	7.23 (4.62)	5.67 (4.41)	6.03 (4.25)	-1.19 (2.23)	5.29 (4.19)
.5	5.06 (3.93)	3.67 (3.52)	3.86 (3.52)	-0.19 (1.97)	2.62 (2.95)
.7	2.89 (2.80)	1.76 (2.24)	2.22 (2.53)	0.11 (1.78)	0.60 (1.29)

Note. 95% confidence intervals were used in the two modified Guttman rules.

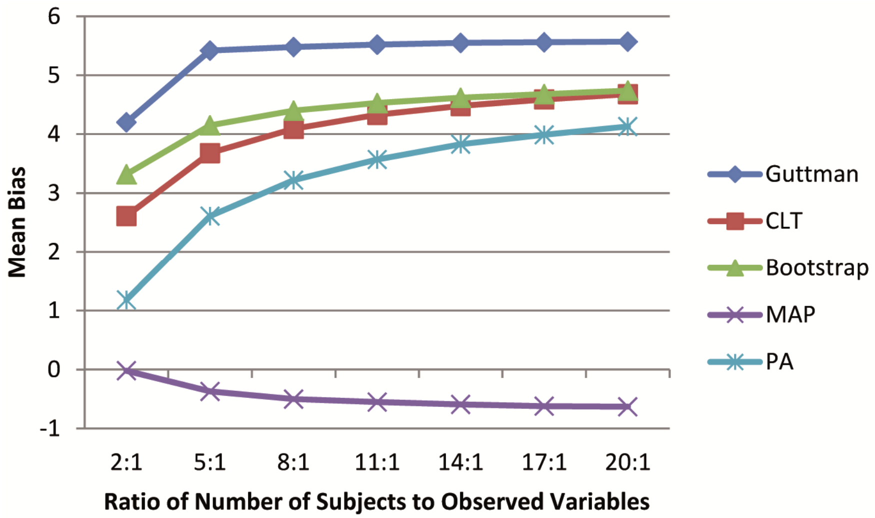

Figure 5:
Line plot of mean bias averaged across the ratios of sample size to observed variables

Discussion

In this Monte Carlo study we studied the bias produced from five different methods of choosing the number of factors in an EFA. In the simulation, we varied the number of observed variables, sample size, true number of factors, factor loading strength, and correlation among factors in order to examine which methods impacted each decision method most strongly. In this section, we will discuss the most important aspects of the study results and the implications for researchers that perform EFAs.

One striking aspect of our results is the consistently poor performance of the Guttman rule across all conditions. Indeed, the Guttman rule was almost always the worst performing method. This poor performance of the Guttman rule is consistent with prior simulation studies on factor decision methods (e.g., Hakstian et al., 1982; Velicer et al., 2000; Zwick & Velicer, 1982, 1986). We believe our study joins the corpus of work that Fabrigar and his colleagues referred to when they stated, ". . . we know of no study of this [Guttman] rule that shows it to work well" (1999, p. 278). The consistent negative findings on the performance of the Guttman rule are particularly distressing when one considers that the Guttman rule remains the most popular method for determining the number of factors in a dataset (Fabrigar et al., 1999; Ford et al., 1986; Hayton et al. 2004; Thompson & Daniel, 1996; Warne et al., 2012). Based on our findings and those from other studies, we recommend that researchers not use the Guttman rule under conditions that resemble those in our simulation.

As for CLT and bootstrap modifications of the Guttman rule, we believe that our previous suggestion to use confidence intervals to help researchers decide the number of factors to retain is an improvement over the prevalent practice in the published research – the original Guttman rule. However, we cannot endorse these modifications above PA or MAP because the modified versions of the Guttman rule suffer from the same weaknesses and shortcomings as the tradition Guttman rule, likely because the modified Guttman methods are minor modifications to the basic flawed premise behind the Guttman rule. The slightly greater robustness that the modified methods demonstrate to the factor loading strength, the number of factors, and the number of observed variables (see Table 2) do not compensate for the problems inherent in the original Guttman rule.

Although the modified Guttman rule is not more accurate than PA or MAP, we still stand by our prior recommendation that researchers report the CIs for their eigenvalues, no matter what factor retention decisions are used. As we stated previously (Larsen & Warne, 2010), we believe that this practice is in accordance with current reporting standards that advocate the use of confidence intervals (e.g., American Psychological Association, 2010; Wilkinson & the Task Force on Statistical Inference, 1999) and also gives readers additional information that permits an informed evaluation of EFA research.

One particularly interesting finding was the impact of sample size on bias. Unsurprisingly, the modified Guttman rules were both susceptible to changes in sample size because larger sample sizes lead to smaller confidence intervals, which in the modified Guttman methods will likely lead to more factors being retained. However, we were surprised to find that PA was a more sensitive method to sample size ($\eta^2 = .12$) than any other method. Indeed, Table 4 shows that as sample size increases, PA tends to overfactor almost as much as the modified Guttman rules. This result has not been previously reported in the simulation literature, likely because previous Monte Carlo studies that investigated the issue had fewer possible simulated sample sizes and thus less ability to detect the sensitivity of PA to variations in sample size. We suggest that future simulation studies be conducted in order to examine the conditions under which PA is sensitive to sample size.

We were also interested in the somewhat contradictory nature of MAP; with a smaller number of observed variables, a larger number of factors in the population, and a stronger correlation among factors, MAP tended to underfactor. MAP's positive performance is due to the fact that it is based on statistical theory, rather than mechanical rules of thumb. Yet, MAP also – for reasons that are unclear to us – can on rare occasions produce a far greater number of factors than exist in a real dataset. Although we encourage readers to use MAP more often than is usually seen in the literature today, we also caution them to keep these caveats in mind, especially because MAP underfactored more than any other method considered. As they use MAP, researchers should be aware that the consequences of underfactoring are generally recognized as being more severe than the consequences of overfactoring (Fava & Velicer, 1992, 1996; Gorsuch, 1983).

There are limitations to this study that the reader should be aware of. First, all factors in the simulation had equal numbers of observed variables. However, it has been shown in previous Monte Carlo simulations that factors that consist of relatively few observed variables compared to the other factors in the population may be difficult to detect (e.g., de Winter, Dodou, & Wieringa, 2009). Second, it is also possible that factors for which all observed variables have the exact same loading is highly unrealistic, although other research has been performed with simulation datasets that have equal factor loadings for each observed variable (e.g., de Winter et al., 2009; Fava & Velicer, 1992, 1996; Zwick & Velicer, 1982). Perhaps a future simulation study could vary the loadings by choosing a mean and standard deviation of loadings to simulate a more realistic set of factor loadings. A third limitation to this study is that there were a fixed number of known true factors, whereas in reality there is sometimes no "true" number of factors. Moreover, researchers may wish to use other criteria – such as the magnitude of factor loadings or the interpretability of the groups of observed variables – to decide how many factors to retain.

Another limitation to our study is the fact that we only investigated the modified Guttman rules with 95% confidence intervals. As the results in the tables show, this confidence interval is not conservative enough to prevent overfactoring. In future research, we hope researchers will examine the viability of other confidence interval widths, such as a 99% interval.

Finally, we urge caution for our readers in applying these results to studies with different conditions than we investigated; there are likely circumstances under which the findings in this article do not hold. We encourage readers to examine other simulation studies about this issue to inform themselves of the factor retention methods that will most likely lead to correct decisions. Also, we hope that readers remember that it is generally recognized as best practice to use multiple methods to determine the appropriate number of factors to retain (Gorsuch, 1983).

Overall, the results of this simulation study show the relationship between the factors varied and the amount of bias produced for each method was fairly predictable. The amount of bias is decreased by: a higher number of factors present in the population data, fewer observed variables, a larger sample size, stronger factor loadings, and a higher correlation among population factors. Also, eigenvalue-based methods are best when there are a relatively small number of observed variables and when the number of factors is large. Moreover, using MAP is a more defensible decision than using eigenvalue-based methods, as MAP is more robust across all conditions in this study. However, MAP does very rarely grossly overfactor, and researchers should be aware of this possibility when using the technique in their analyses.

We believe that this study provides valuable information for future data analysts as they select the appropriate factor retention method(s). The study may also help researchers understand the conflicting results that different factor retention methods can produce when used in conjunction with one another. We hope this study aids in the understanding of factor retention methods and the causes of potential bias among them. Finally, we join with previous researchers on factor retention (e.g., Fabrigar et al., 1999; Velicer et al., 2000) in calling for manufacturers and programmers to remove the Guttman rule from the default factor retention setting in their statistical packages and to replace it with a more accurate method. We believe that the decades of consistent findings are doing little to lessen the prominence of the Guttman rule in the published research and that only changes in the default settings of software will spawn a majority of researchers to make better factor retention decisions.

Author note

Russell T. Warne, Department of Behavioral Science, Utah Valley University; Ross Larsen, Curry School of Education, University of Virginia.

Ross Larsen is now at the Department of Foundations of Education, School of Education, Virginia Commonwealth University.

The authors thank Rosalma Arcelay for her feedback on earlier drafts of the manuscript.

References

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin, 103*, 276-279. doi:10.1037/0033-2909.103.2.276

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

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

Cota, A. A., Longman, R. S., Holden, R. R., Fekken, G. C., & Xinaris, S. (1993). Interpolating 95th percentile eigenvalues from random data: An empirical example. *Educational and Psychological Measurement, 53*, 585-596. doi:10.1177/0013164493053003001

de Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*, 147-181. doi:10.1080/00273170902794206

Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment, 27*, 494-507. doi:10.1177/0734282909333179

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27*, 387-415. doi:10.1207/s15327906mbr2703_5

Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analyses. *Educational and Psychological Measurement, 56*, 907-929. doi:10.1177/0013164496056006001

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

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55*, 377-393. doi:10.1177/0013164495055003002

Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 143-164). Hoboken, NJ: Wiley.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*, 149-161. doi:10.1007/BF02289162

Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research, 17*, 193-219. doi:10.1207/s15327906mbr1702_3

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

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393-416. doi:10.1177/0013164405282485

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

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151. doi:10.1177/001316446002000116

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Larsen, R., & Warne, R. T. (2010). Estimating confidence intervals for eigenvalues in exploratory factor analysis. *Behavior Research Methods, 42*, 871-876. doi:10.3758/BRM.42.3.871

Liu, O. L., & Rijmen, F. (2008). A modified procedure for parallel analysis for ordered categorical data. *Behavior Research Methods, 40*, 556-562. doi:10.3758/brm.40.2.556

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396-402. doi:10.3758/BF03200807

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517-529. doi:10.1037//0021-9010.873.3.517

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*, 25-36. doi:10.1177/0146621697211002

SAS. (2009). SAS/STAT 9.2 user's guide (2nd ed.). Cary, NC: SAS Institute.

Shönemann, P. H. (1990). Facts, fictions, and common sense about factors and components. *Multivariate Behavioral Research, 25*, 47-51. doi:10.1207/s15327906mbe2501_5

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208. doi:10.1177/0013164496056002001

Tonsor, S. (2004, December 7). PCA bootstrap confidence intervals program [SAS code]. Retrieved from http://www.pitt.edu/~tonsor/downloads/programs/bootpca.html

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

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41-71). New York, NY: Kluwer Academic/Plenum Publishers.

Warne, R. T., Lazo, M., Ramos, T., & Ritter, N. (2012). Statistical methods used in gifted education journals, 2006-2010. *Gifted Child Quarterly, 56*, 134-149. doi:10.1177/0016986212444122

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604. doi:10.1037/0003-006X.54.8.594

Zheng, B., Hall, M. A., Dugan, E., Kidd, K. E., & Levine, D. (2002). Development of a scale to measure patients' trust in health insurers. *Health Services Research, 37*, 187-202.

Zientek, L. R., & Thompson, B. (2007). Applying the bootstrap to the multivariate case: Bootstrap component/factor analysis. *Behavior Research Methods, 39*, 318-325. doi:10.3758/BF03193163

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17*, 253-269. doi:10.1207/s15327906mbr1702_5

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442. doi:10.1037/0033-2909.99.3.432