Relative Accuracy of Two Modified Parallel Analysis Methods that Use the Proper Reference Distribution

Educational and Psychological
Measurement

2018, Vol. 78(4) 589–604

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164417718610

journals.sagepub.com/home/epm

Samuel Green¹, Yuning Xu¹, and Marilyn S. Thompson¹

Abstract

Parallel analysis (PA) assesses the number of factors in exploratory factor analysis. Traditionally PA compares the eigenvalues for a sample correlation matrix with the eigenvalues for correlation matrices for 100 comparison datasets generated such that the variables are independent, but this approach uses the wrong reference distribution. The proper reference distribution of eigenvalues assesses the k th factor based on comparison datasets with $k-1$ underlying factors. Two methods that use the proper reference distribution are revised PA (R-PA) and the comparison data method (CDM). We compare the accuracies of these methods using Monte Carlo methods by manipulating the factor structure, factor loadings, factor correlations, and number of observations. In the 17 conditions in which CDM was more accurate than R-PA, both methods evidenced high accuracies (i.e., $> 94.5\%$). In these conditions, CDM had slightly higher accuracies (mean difference of 1.6%). In contrast, in the remaining 25 conditions, R-PA evidenced higher accuracies (mean difference of 12.1%, and considerably higher for some conditions). We consider these findings in conjunction with previous research investigating PA methods and concluded that R-PA tends to offer somewhat stronger results. Nevertheless, further research is required. Given that both CDM and R-PA involve hypothesis testing, we argue that future research should explore effect size statistics to augment these methods.

Keywords

exploratory factor analysis, parallel analysis, psychometrics

¹Arizona State University, Tempe, AZ, USA

Corresponding Author:

Samuel Green, School of Social and Family Dynamics, Arizona State University, PO Box 873701, Tempe, AZ 85287-370, USA.

Email: samgreen@asu.edu

Based on the psychometric literature, a number of factor analytic experts have recommended the use of parallel analysis (PA) to assess the number of factors underlying a set of measures when conducting an exploratory factor analysis (EFA; e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Preacher & MacCallum, 2003). As portrayed in the first column of Table 1, traditional parallel analysis (T-PA) involves the following steps: (a) Conduct a principal component analysis (PCA) on sample data and record the eigenvalues for the sequential components. (b) Generate 100 or more comparison datasets with the same number of variables and sample size as the sample data. The scores for the comparison datasets are generated assuming the variables are multivariate normally distributed and uncorrelated in the population. (c) Conduct PCAs on the comparison datasets to create reference distributions of eigenvalues for the sequential components. (d) Calculate the mean eigenvalue for each sequential component. (e) Determine sequentially the first component that has a sample eigenvalue which fails to exceed the respective mean of eigenvalues for the comparative datasets. Given that it is the k th sequential component, the estimated number of factors is $k-1$.

Despite the accuracy of T-PA relative to other methods for assessing the number of factors, Harshman and Reddon (1983) and Turner (1998) believed T-PA was flawed. They argued against the use of a reference distribution of eigenvalues for the comparison datasets that assumes the variables are uncorrelated. They contended the proper reference distribution of eigenvalues to reach a conclusion about the k th factor should be based on datasets with $k-1$ underlying factors. Within this perspective, the use of a reference distribution of eigenvalues based on uncorrelated variables is appropriate to reach conclusions for only the first factor.

To address concerns raised by Harshman and Reddon (1983) as well as Turner (1998), Green, Levy, Thompson, Lu, and Lo (2012) and Ruscio and Roche (2012) independently suggested revisions in conducting PA. They referred to their approaches as revised parallel analysis (R-PA) and the comparison data method (CDM), respectively. When assessing whether at least k factors underlie a set of measures with R-PA and CDM, comparison datasets are generated assuming measures are a function of $k-1$ factors rather than 0 factors, as with T-PA. Ideally, the comparison datasets should be generated based on the population loadings of these $k-1$ factors; however, the population factor loadings are unknown in practice. Accordingly, with these methods, factor loadings are based on the sample dataset. Results of Monte Carlo studies offer support for R-PA and CDM (Green et al., 2012; Green, Redell, Thompson, & Levy, 2016; Green, Thompson, Levy, & Lo, 2015; Ruscio & Roche, 2012; Socha & Bandalos, 2015), although neither method yields uniformly greater accuracy relative to other parallel analysis methods across all conditions.

Although R-PA and CDM use similar methods to generate comparative datasets, the approaches differ from each other in a number of ways. In the last two columns of Table 1, we delineate the steps required by R-PA and CDM, showing that R-PA and CDM differ from each other at every step. Given both have been found to yield

Table 1. The Six Steps to Assess the Need for a *k*th Factor With T-PA, R-PA, and CDM.

T-PA	R-PA	CDM
<i>Analysis of sample data</i>		
1. Compute eigenvalues and eigenvectors of correlation matrix for sample data with <i>p</i> variables (principal components).	1. Compute eigenvalues and eigenvectors of reduced correlation matrix (R^2 on diagonal) for sample data with <i>p</i> variables (principal axis).	1. Compute eigenvalues and eigenvectors of correlation matrix for sample data with <i>p</i> variables (principal components).
<i>Generation and analysis of comparison datasets (CDs)</i>		
2. Generate 100 CDs assuming the measures are uncorrelated.	2. Generate 100 CDs based on sample loadings for the <i>k</i> – 1 factors of the sample data using principal axis factoring.	2. Draw 500 CDs from a population of 10,000, generated so that the population correlation matrix is a function of the loadings for <i>k</i> – 1 factors and is as similar as possible to the sample correlation matrix.
3. Compute eigenvalues and eigenvectors for the correlation matrices for the 100 CDs (principal components).	3. Compute eigenvalues and eigenvectors for the reduced correlation matrices (R^2 on diagonal) for the 100 CDs (principal axis).	3. Compute eigenvalues and eigenvectors for the correlation matrices for the 500 CDs (principal components).
4. Compute mean of the eigenvalues for the <i>k</i> th factor of the CDs.	4. Compute 95th percentile of the eigenvalues for the <i>k</i> th factor of the CDs.	4. Compute RMSRs between the <i>p</i> eigenvalues for the sample dataset and the <i>p</i> eigenvalues for CDs generated given <i>k</i> – 1 factors.
5. Compare the <i>k</i> th eigenvalue for the sample data to the mean of the eigenvalues for the <i>k</i> th factor of the 100 CDs.	5. Compare the <i>k</i> th eigenvalue for the sample data to the 95th percentile of the eigenvalues for the <i>k</i> th factor of the 100 CDs.	5. Repeat Steps 2-4 assuming <i>k</i> factors. Conduct a one-tailed Mann–Whitney <i>U</i> test ($\alpha = .30$) to compare the 500 RMSRs for CDs generated given <i>k</i> – 1 factors and the 500 RMSRs for CDs generated given <i>k</i> factors.
<i>Decision rule to assess number of factors</i>		
6. If the <i>k</i> th eigenvalue for the sample is greater than the mean eigenvalue, at least <i>k</i> factors underlie the data and proceed to the next step. If not, stop the stepwise process and the assessed number of factors is equal to <i>k</i> – 1.	6. If the <i>k</i> th eigenvalue for the sample is greater than the 95th percentile of the eigenvalues for the CDs and greater than 0, at least <i>k</i> factors underlie the data and proceed to the next step. If not, stop the stepwise process and the assessed number of factors is equal to <i>k</i> – 1.	6. If rejected, conclude that at least <i>k</i> factors underlie the data and proceed to the next step. If not rejected, stop the stepwise process and the assessed number of factors is equal to <i>k</i> – 1.

Note. T-PA = traditional parallel analysis; R-PA = revised parallel analysis; CDM = comparison data method; RMSR = root mean square residuals.

relatively accurate results, but differ in many ways from each other, we chose to conduct a study to compare the accuracy of these two methods using Monte Carlo methods. Before describing our study more fully, we discuss in greater detail the R-PA and CDM approaches, paying particular attention to their differences.

Revised Parallel Analysis

We briefly describe the modifications to T-PA applied in the creation of R-PA. We then review Monte Carlo studies that evaluated the accuracy of R-PA to estimate the number of factors underlying a set of measures.

As shown in the first two columns of Table 1, R-PA differs from T-PA in two ways in addition to the choice of reference distributions of eigenvalues. One difference is that R-PA uses principal axis factoring (PAF) rather than principal components. The use of common factor analysis, such as PAF, is preferred to PCA by many psychometricians in evaluating the underlying structure of measures because the common factor analytic model allows for measures with less than perfect reliability, a ubiquitous phenomenon in the social sciences. From this perspective, parallel analysis should be conducted with a common factor analytic method to allow for measurement error (e.g., Ford, MacCallum, & Tait, 1986). A second difference is that the eigenvalues for the sample data are compared with the 95th percentile of eigenvalues for the referent distribution with R-PA rather than the mean eigenvalue as with T-PA. This modification follows recommendations by researchers who endorsed a more stringent rule in assessing the number of factors (e.g., Buja & Eyuboglu, 1992; Glorfeld, 1995). In addition, the use of the 95th percentile eigenvalue rule in conducting parallel analysis makes it consistent with traditional applications of hypothesis testing in which the probability of a Type I error is set at .05.

Crawford et al. (2010) investigated the accuracy of PA using PAF versus PCA and the mean eigenvalue rule versus 95th percentile eigenvalue rule. PA was conducted with reference distributions of eigenvalues based on uncorrelated measures. They found that the relative accuracies of the different methods were dependent on the data generation conditions (e.g., sample size, number of factors, and magnitude of the correlations between factors). Even more important, the various methods did not behave well statistically; that is, their accuracies failed to increase consistently as sample size, factor loadings, and number of variables per factor increased and as correlations between factors decreased.

In a follow-up study, Green et al. (2012) introduced the use of a reference distribution of eigenvalues based on $k-1$ factors. In this study, they compared the accuracy of PA involving combinations of the following approaches: (a) PCA or PAF, (b) the mean or the 95th percentile eigenvalue rule, and (c) the reference distribution of eigenvalues based on 0 factors (i.e., uncorrelated measures) or reference distribution of eigenvalues based on $k-1$ factors. R-PA using PAF, the 95th percentile rule, and a reference distribution based on $k-1$ factors yielded relatively accurate results and behaved better statistically than the other methods. PA using PAF, the 95th

percentile rule, and reference eigenvalue distribution based on uncorrelated measures also demonstrated relatively high accuracy, but did not behave as well statistically. We refer to this latter method as modified traditional PA (MT-PA); it differs from R-PA only with respect to the choice of a reference distribution.

Subsequently, Green, Thompson, Levy, and Lo (2015) assessed the accuracy of R-PA and MT-PA within the framework of hypothesis testing. Because the two PA methods use the 95% percentile rule, they can be viewed as a series of hypothesis tests with alpha set to .05. In addition, Green et al. (2015) included traditional likelihood ratio tests (LRTs; Hayashi, Bentler, & Yuan, 2007) to estimate the number of factors. In terms of overall accuracy, both PA approaches generally outperformed the LRT method. In comparing the two PA methods, MT-PA tended to be more accurate in conditions with low factor loadings, whereas R-PA was more accurate for conditions with high correlations between factors, conditions with an underlying model that included both general and group factors (i.e., bifactor models), and conditions in which some of the variables were not a function of any common factors. From a hypothesis testing perspective, the alphas for the MT-PA approach tended to be too conservative in conditions with high factor loadings, particularly with larger sample sizes, and too liberal in conditions with low factor loadings and greater numbers of factors. In contrast, the alphas for R-PA were below or close to .05 in all conditions except one. The alphas tended to be overly conservative with high factor loadings and with more than a single underlying factor. In conditions with lower factor loadings, MT-PA had greater power than the revised method, which produced greater accuracy; however, this greater power was due to inflated alphas. In contrast, R-PA showed greater accuracy under a variety of conditions due to its greater power rather than inflated alphas.

In a related study, Green et al. (2016) evaluated the relative accuracy of R-PA and MT-PA methods for factor analysis of tetrachoric correlations between items with binary responses. The results were similar to those for Green et al. (2015) except the accuracies tended to be lower overall.

Comparison Data Method

Concurrent with the work by Green and his colleagues, Ruscio and Roche (2012) proposed CDM. Like R-PA, CDM is a sequential approach that uses an appropriate statistical reference distribution. With CDM (as described in Table 1), principal component analysis is conducted on the sample correlation matrix to obtain eigenvalues and eigenvectors. These eigenvalues subsequently are compared with the eigenvalues extracted based on the comparison datasets. To create comparison datasets, 10,000 cases are generated using the GenData program (Ruscio & Kaczetow, 2008) such that the correlation matrix associated with these data can be reproduced by $k-1$ underlying factors and is as similar as possible to the sample correlation matrix. Five-hundred random samples are then drawn from this dataset, principal components analysis is applied, and eigenvalues are computed. Root mean square residuals

(RMSRs) are calculated between the eigenvalues for all components of each comparative dataset and the eigenvalues for all components of the sample dataset. An identical process is conducted to obtain RMSRs for comparative datasets based on k underlying factors. The 500 RMSRs based on comparative datasets with $k-1$ underlying factors are compared with the 500 RMSRs based on comparative datasets with k underlying factors. The decision rule for the number of factors is determined by a series of comparisons of the RMSRs for k factors and RMSRs with $k-1$ factors using the Mann–Whitney U test (M-W test) with alpha set at .30. In the earliest step in which the null hypothesis is not rejected, the number of estimated factors is equal to $k-1$, as defined at that step.

Ruscio and Roche (2012) conducted a Monte Carlo study to assess the accuracy of CDM against eight other methods, including PA using comparison datasets assuming uncorrelated measures. Generally, they found their proposed method produced more accurate estimates of the number of factors in comparison with traditional PA as well as the other approaches. In contrast, Socha and Bandalos (2015) found that traditional PA with the 95% eigenvalue rule (rather than the mean eigenvalue rule) tended to be more accurate than CDM with less complex models. However, neither the studies by Ruscio and Roche (2012) nor by Socha and Bandalos (2015) included R-PA as an alternative method to assess the number of factors.

With CDM, all eigenvalues for the comparative datasets are compared with the eigenvalues of sample datasets by computing RMSRs between them. Accordingly, relative to other PA approaches that focus on an eigenvalue for a particular sequential component (or factor), CDM is a more comprehensive assessment of the fit of a model to the sample data. Although not explicitly stated by Ruscio and Roche (2012), we assume they chose to use principal components analysis to minimize complexities associated with computing RMSRs across all eigenvalues. In particular, it is unclear how to adapt CDM if principal axes factoring is used because PAF yields negative eigenvalues, and the number of negative eigenvalues may differ across comparative datasets.

The assessment of all eigenvalues also required an alternative rule to assess the estimated number of factors. They use the M-W test, and set alpha equal to .30 because it appeared to be optimal based on a series of Monte Carlo analyses. This heuristic approach yielded relatively accurate results based on their Monte Carlo study. It should be noted, however, that one would expect that the CDM method to have an upper-bound accuracy of .70 given the null hypothesis for the M-W test holds and alpha is set to .30. In contrast, they reported accuracies that tended to range between 85% and 95%.

Objective of Research

The primary purpose of our Monte Carlo study was to evaluate the relative accuracies of CDM and R-PA to estimate the number of factors. We evaluated these methods across conditions that varied with respect to sample size and the specification of

the factor model underlying the generated data. When the methods yielded inaccurate estimates, we noted whether the number of factors was underestimated or overestimated.

Previously, Green et al. (2015) compared the accuracy of R-PA and MT-PA. We chose to design the current study to be very similar to this previous study to allow us to assess the relative accuracies of not only CDM with R-PA, but also CDM with MT-PA. Given R-PA was more accurate than MT-PA under some, but not all, conditions in the Green et al. (2015) study, the evaluation of the relative accuracy of CDM with MT-PA based on the two studies would not be possible without a common methodology.

Method

Design

We manipulated four data generation dimensions: type of factor model, factor loadings, factor correlations, and sample size. These generation dimensions are described next.

- *Type of factor models:* Data were generated with seven types of factor models. Models with two or fewer factors were specified with 8 items, whereas models with three factors were specified with 12 items. The different types of generation models were (a) a baseline, zero-factor model in which all eight items were a function of only error; (b) a one-factor model in which all eight items loaded on a single factor; (c) a one-factor model in which four items loaded on a single factor, and the other four items were a function of only error (referred to as a one-factor model with unique items); (d) a two-factor, perfect-clusters model, with four items loading on each of the two factors; (e) a three-factor, perfect-clusters model, with four items loading on each of the three factors; (f) a two-factor, bifactor model, with all 8 items loading on a general factor and 4 items also loading on a group factor; and (g) a three-factor bifactor model, with all 12 items loading on a general factor, 4 of the 12 items loading on one group factor, and another 4 items loading on a second group factor.
- *Factor loadings:* For the one-factor models, loadings for all items were either .5s or .7s. For the one-factor model with unique items, loadings on the factor were either .5s or .7s for four items and 0s for the four remaining items. For two-factor or three-factor, perfect-clusters models, the nonzero loadings on the factors were either all .5s or all .7s. For the bifactor models, the items on the general factor had loadings of either all .5s or all .7s, and the 4 items on the group factor(s) had loadings of .5s.
- *Factor correlations:* For any two-factor or three-factor, perfect-clusters model, all correlations between factors were 0, .5, or .8. For the bifactor models, the correlation between factors was always 0.
- *Number of observations:* The number of observations was 200 or 400.

We included bifactor models to generate data in that they have been found to fit data from a wide variety of educational and psychological research studies (Reise, 2012). In practice, prior to conducting an EFA bifactor rotation (Jennrich & Bentler, 2011; Reise, Moore, & Maydeu-Olivares, 2011), it is crucial that parallel analysis be able to estimate accurately the number of factors, including group factors as well as the general factor.

Data Generation and Analyses

For each of the 42 data generation conditions, 1,000 sample datasets were generated with a common factor model. The factors and the errors in the model were generated to be normally distributed. Correlation matrices were computed for each sample dataset and analyzed using R-PA and CDM as described in Table 1.

For R-PA, we generated 100 comparative datasets for each sample dataset, and used PAF to analyze both the sample and comparative datasets. We applied the 95th percentile eigenvalue rule to the sampling distributions of eigenvalues to assess the number of underlying factors. An *R* program for conducting R-PA is available at <https://web.asu.edu/samgreen/software>.

CDM was implemented by the *R* program available from Ruscio and Roche (2012). This *R* program generates 500 comparative datasets for each sample dataset and applies PCA to analyze both the sample and comparative datasets. RMSRs are computed between the sample eigenvalues and the eigenvalues for each of the comparative datasets. A one-tailed M-W test is conducted at $\alpha = .30$ between the 500 RMSRs based on comparative datasets with k factors and the 500 RMSRs for comparative datasets with $k-1$ factors. The program was designed to assess whether one or more factors underlie sample data. We extended the program to also consider 0 factors.

For each condition, we computed the percentage of sample datasets in which R-PA and CDM accurately estimated the number of factors, underestimated the number of factors, and overestimated the number of factors.

Results

We report the results for conditions with 0 or 1 factor models in Table 2, models with 2 or 3 perfect-cluster factors in Table 3, and bifactor models with two factors or three factors in Table 4. In these tables, we present the accuracies as well as the percentages of cases in which the estimated number of factors was underpredicted or overpredicted.

In all six conditions with models having 0 or 1 factor underlying all items, both CDM and R-PA had accuracies of 94.7% or greater. CDM had slightly higher accuracies in these conditions, with the largest difference being 4.1%. In contrast, for all four conditions with 1 factor but with half the items not loading on a factor, the accuracies were slightly greater for R-PA, with the largest difference being 5.0%.

Table 2. Accuracies (%) of CDM and R-PA for Models With 0 or 1 Factor.

		CDM			R-PA		
λ	N	Accuracy	Underpredict	Overpredict	Accuracy	Underpredict	Overpredict
Model with no underlying factors							
—	200	96.1	—	3.9	95.0	—	5.0
	400	95.3	0.0	4.7	94.7	—	5.3
One-factor model for all items							
.5	200	98.5	0.0	1.5	95.9	0.0	4.1
	400	99.1	0.0	0.9	97.5	0.0	2.5
.7	200	98.7	0.0	1.3	95.6	0.0	4.4
	400	99.3	0.0	0.7	95.2	0.0	4.8
One-factor model with unique items							
.5	200	92.3	0.0	7.7	94.2	0.0	5.8
	400	92.8	0.0	7.2	95.0	0.0	5.0
.7	200	89.7	0.0	10.3	94.7	0.0	5.3
	400	89.6	0.0	10.4	94.1	0.0	5.9

Note. CDM = comparison data method; R-PA = revised parallel analysis.

In models with perfect-cluster factors (as shown in Table 3), CDM was more accurate than R-PA in 11 of the 24 conditions. In all these conditions, both CDM and R-PA had accuracies of 96.0% or greater; CDM had slightly greater accuracies, with the largest difference being 2.0%. Conditions in which CDM was slightly greater tended to be those with smaller correlations between factors. In the other 13 conditions in Table 3, R-PA had greater accuracies, and in some conditions substantially greater. For example, the accuracies for R-PA exceeded those for CDM by 10% or more in 5 conditions; four of these conditions were those in which the correlation between factors was .8. For conditions with factor correlations of .8 (except those with loadings of .7 and sample size of 400), errors in determining the number of factors generally underpredicted the number of factors with R-PA, but to a lesser degree than with CDM. It should be noted that both methods were quite inaccurate (i.e., less than 50% accuracies) when the factors were correlated .8 and had loadings of .5.

R-PA produced higher accuracies in all conditions with bifactor models (as presented in Table 4). The differences were substantial; the smallest difference was 12.7%, and the largest difference was 34.3%. When accuracies were lower (e.g., less than 60%), the inaccuracies were generally due to underprediction of the number of factors.

Integrating Results of the Current Study With a 2015 Study Comparing MT-PA and R-PA

In a previous study (Green et al., 2015; for simplicity, referred to here as the 2015 study), the accuracies of R-PA and MT-PA were compared under the same

Table 3. Accuracies (%) of CDM and R-PA for Models With Two or Three Perfect-Cluster Factors.

Correlation between factors			CDM			R-PA		
λ	N		Accuracy	Underpredict	Overpredict	Accuracy	Underpredict	Overpredict
Model with two perfect-cluster factors								
0	.5	200	98.3	0.0	1.7	97.7	0.0	2.3
		400	98.9	0.0	1.1	97.3	0.0	2.7
	.7	200	97.8	0.0	2.2	96.0	0.0	4.0
		400	97.9	0.0	2.1	96.5	0.0	3.5
.5	.5	200	85.2	11.9	2.9	94.6	0.0	5.4
		400	98.6	0.2	1.2	96.6	0.1	3.3
	.7	200	98.3	0.0	1.7	96.5	0.0	3.5
		400	97.5	0.0	2.5	97.2	0.0	2.8
.8	.5	200	11.0	88.8	0.2	22.9	76.2	0.9
		400	31.5	67.7	0.8	49.5	46.9	3.6
	.7	200	89.4	8.2	2.4	95.1	1.9	3.0
		400	97.1	0.0	2.9	97.2	0.0	2.8
Model with three perfect-cluster factors								
0	.5	200	97.6	0.1	2.3	96.1	0.0	3.9
		400	98.8	0.0	1.2	97.7	0.0	2.3
	.7	200	94.8	0.0	5.2	96.3	0.0	3.7
		400	96.2	0.0	3.8	97.8	0.0	2.2
.5	.5	200	50.1	48.8	1.1	69.0	27.5	3.5
		400	95.8	3.1	1.1	97.3	0.1	2.6
	.7	200	97.5	0.0	2.5	96.5	0.0	3.5
		400	98.6	0.0	1.4	96.8	0.0	3.2
.8	.5	200	0.7	97.3	0.0	3.1	96.7	0.2
		400	2.2	97.8	0.0	14.3	84.5	1.2
	.7	200	57.1	41.3	1.6	84.5	12.6	2.9
		400	96.7	1.6	1.7	97.7	0.0	2.3

Note. CDM = comparison data method; R-PA = revised parallel analysis.

conditions investigated in the current study, except no conditions were included investigating zero factors. It is relevant and interesting to integrate the results of the present study with those of the 2015 study to make comparisons of accuracies among R-PA, CDM, and MT-PA. We note that the R-PA computer programs for the two studies were written by different programmers using different programming languages (SAS vs. R), thus invoking different subroutines for generating data and conducting factor analyses.

Before making comparisons across different methods, we assessed the comparability of results for R-PA across the two studies. For all conditions, the mean absolute difference in accuracies for R-PA was 1.9%, with no difference for a condition exceeding 5%. The accuracies in the 2015 study were slightly higher, with an overall mean difference of 0.6%. The differences varied little as a function of the 6 factor

Table 4. Accuracies (%) of CDM and R-PA for Bifactor Models.

		CDM			R-PA		
λ	N	Accuracy	Underpredict	Overpredict	Accuracy	Underpredict	Overpredict
Bifactor model with two factors							
.5	200	57.8	35.9	6.3	80.0	16.0	4.0
	400	83.7	5.9	10.4	96.4	0.5	3.1
.7	200	80.3	2.8	16.9	98.6	0.3	2.9
	400	81.7	0.0	18.3	97.4	0.0	2.6
Bifactor model with three factors							
.5	200	27.0	71.5	1.5	43.1	54.4	2.5
	400	50.3	43.6	6.1	81.0	15.7	3.3
.7	200	55.7	36.8	7.5	90.0	6.7	3.3
	400	68.9	11.3	19.8	95.5	1.5	3.0

Note. CDM = comparison data method; R-PA = revised parallel analysis.

model types. The largest mean difference for a factor model type was 1.6% for the bifactor model with one group factor; this mean difference was higher for the 2015 study. The second largest mean difference for a factor model type was 1.0% for the bifactor model with two group factors; this mean difference was greater for the current study. Overall, we viewed the differences in R-PA accuracies between the two studies as relatively small.

Next, we compared the MT-PA accuracies based on the 2015 study with the R-PA accuracies based on the current study. In 17 of the 20 conditions in which both MT-PA and R-PA had accuracies of 94.5% or greater, MT-PA was slightly more accurate than R-PA. The mean difference for all 20 of these conditions was 2.6% in favor of MT-PA, with largest difference being 4.8%. In the remaining 20 conditions, R-PA outperformed MT-PA; the accuracies in these conditions varied dramatically from 3.1% to 98.6% for R-PA and from 0.7% to 93.5% for MT-PA. The mean difference in these 20 conditions was 20.0% in favor of R-PA, with the largest difference being 74.5%. Differences were generally greatest for models with highly correlated factors and bifactor models with two group factors, particularly with the smaller sample size of 200. Consistent with the conclusions reached in the 2015 study, R-PA overall yielded more positive results than MT-PA for the conditions explored.

Finally, we compared the MT-PA accuracies for the 2015 study with the CDM accuracies for the current study. In 17 of the 19 conditions in which both MT-PA and CDM had accuracies of 94.5% or greater, MT-PA was more accurate than CDM. The mean difference for these 19 conditions was 1.7% in favor of MT-PA, with the largest difference being 5.2%. In the remaining 21 conditions, MT-PA outperformed CDM in 9 conditions, and CDM outperformed MT-PA in 10 conditions; accuracies were equivalent in the other 2 conditions. In the conditions in which MT-PA had higher accuracies, the differences in accuracies ranged from 0.3% to 18.0%, with a mean accuracy of 6.0%. In the conditions in which CDM had higher accuracies, the

differences in accuracies ranged from 1.0% to 40.0%, with a mean accuracy of 15.4%. Differences in favor of CDM tended to be greatest for models with highly correlated factors and bifactor models with two group factors. Based on these results, CDM was more accurate in some conditions, whereas MT-PA was more accurate in others. However, in conditions in which CDM yielded better results in comparison with the conditions in which MT-PA performed better, the differences in accuracies tended to be greater.

Discussion

Choice Among Methods to Decide on the Number of Factors

In the current study, R-PA outperformed CDM in 25 of the 42 conditions. The differences in accuracy were on average 12.1% higher for R-PA in comparison with CDM, with the largest difference being 34.3%. In these conditions, the accuracies varied dramatically from 3.1% to 98.6% for R-PA and from 0.7% to 97.1% for CDM. On the other hand, CDM was slightly more accurate than R-PA in assessing the number of factors in the remaining 17 conditions. However, in these conditions, both CDM and R-PA had accuracies of 94.5% or greater; the largest difference in favor of CDM was 4.1%, and the mean difference in conditions favoring CDM was 1.6%. Because neither method was uniformly more accurate across conditions, we cannot make an absolute recommendation. However, given these results, we suspect most applied researchers would prefer R-PA because it performed nearly as well as CDM when CDM was more accurate and R-PA was substantially more accurate in other conditions. Additionally, R-PA was less prone to underfactoring than CDM; underfactoring is regarded as more problematic than overfactoring in that an underfactored solution can obscure the factor pattern and omit key factors (Fabrigar et al., 1999).

R-PA outperformed CDM in conditions with the following types of models: one-factor model with unique items; two-factor, perfect-clusters models with more highly correlated factors; and bifactor models. It is not surprising that the latter two types of models produced similar results. The two-factor, perfect-clusters models with more highly correlated factors are equivalent to hierarchical models with a second-order general factor, which is nested within a bifactor model (Rindskopf & Rose, 1988). Thus, we can conclude that R-PA detected group factors in the presence of a general factor more accurately than CDM.

By comparing the results from the current study with those from the 2015 study by Green et al., we also were able to assess the relative accuracy of MT-PA with the accuracies of R-PA and CDM. Based on these two studies, we concluded that R-PA performed nearly as well as MT-PA when both methods evidenced high accuracy. On the other hand, R-PA was more accurate in the remaining conditions and substantially so in some cases. This pattern of results is very similar to the one found in comparing the accuracies of R-PA and MT-PA in Green et al.'s (2015) study. When comparing the accuracies of MT-PA and CDM, CDM was more accurate in some conditions, whereas MT-PA was more accurate in others. Overall, the preferred PA

method appeared to be R-PA relative to the other two methods, at least for the conditions explored.

When the generation models contained more correlated factors or had a bifactor structure, R-PA tended to evidence higher accuracies relative to CDM in the current study and relative to MT-PA in the Green et al. (2015) study. As previously discussed, CDM and MT-PA apply principal components analysis, whereas R-PA uses principal axes factor analyses. Accordingly, it seems reasonable to hypothesize that the choice of extraction methods played a major role in the pattern of results for the two studies.

Future Research

Many studies have evaluated the relative effectiveness of PA and alternative methods for assessing the number of factors as well as the relative accuracy of different PA modifications. On the surface, the conclusion might be that there is little to gain in conducting additional research involving PA. In this section, we argue that more research is crucial to apply PA methods optimally, and we suggest a number of areas that might be investigated in the future.

Assessment of Differences in Methods for R-PA and CDM. As shown in Table 1, R-PA and CDM differ in a number of ways. For example, R-PA involves principal axes factor analysis, the eigenvalue for the k th factor for the sample data, a reference distribution of eigenvalues for the k th factor based on the 100 comparative datasets, and a hypothesis test based on the reference distribution at the .05 level. In contrast, CDM involves principal component analysis, eigenvalues for all components for the sample data, RMSRs between the eigenvalues for the sample dataset and the eigenvalues for each of the 500 comparative datasets, and a one-tailed M-W test comparing the 500 RMSRs given $k-1$ factors and the 500 RMSRs given k factors.

The major innovation with CDM is the computation of RMSRs between all eigenvalues for the sample data and a comparative dataset. Many of the other differences with CDM were likely invoked to take into account the use of RMSRs. For example, as previously discussed, it is unclear how to apply CDM using principal axes factoring rather than principal components because PAF yields negative eigenvalues, and the number of negative eigenvalues may differ across comparative datasets. Nevertheless, future research could consider whether changes to particular components of R-PA and CDM would produce changes in their relative accuracies. To illustrate, we considered the effect of alpha levels for R-PA and CDM on their relative accuracies. We discuss this line of future research in the next section.

Choice of Alphas for R-PA and CDM. To assess the utility of varying alpha levels for R-PA and CDM in future research, we conducted a limited investigation in which we examined two conditions. In both conditions, we increased the alphas by .15 for the two approaches: .05 to .20 for R-PA and .30 to .45 for CDM. In the first condition,

we generated data using a bifactor model with two group factors, factor loadings of .5 on the general factor, and a sample size of 200. In the current study, both R-PA with $\alpha = .05$ and CDM with $\alpha = .30$ badly underestimated the correct number of factors for this condition. We expected that increasing their alpha levels would decrease underestimation and yield greater accuracy. In the other condition, data were generated using a one-factor model with unique items, factor loadings of .5, and a sample size of 200. Both R-PA and CDM had relatively high accuracies for this condition; inaccuracy was due to overestimation of the number of factors. We hypothesized that liberalization of the stopping rule for these PA methods (i.e., raising the α s) would increase overestimation of the correct number of factors and thus decrease accuracy.

The results for the bifactor model were as expected. The accuracies for CDM increased from 27.0% to 43.8% when α was increased from .30 to .45. This change was due to a decrease in underestimation: 71.5% to 44.5%. The accuracies for R-PA increased from 43.1% to 58.0% when α was increased from .05 to .20. The improved accuracy was due to the decrease in underestimation from 54.4% to 28.6% with an increase in α s from .05 to .20. However, the change in α s with R-PA also increased overestimation: 2.5% to 13.4%.

For the one-factor model with unique items, the accuracies for CDM decreased slightly from 92.3% to 91.5% when α was increased. The change was totally due to an increase in overestimation of the number of correct factors (7.7% to 8.5%). In comparison, accuracies for R-PA decreased markedly from 94.2% to 77.4% when α was increased from .05 to .20. Again, the change was totally due to an increase in overestimation of the number of the correct factors (5.8% to 22.6%).

As expected, the number of estimated factors using R-PA and CDM tended to increase when α s were increased by .15. For the bifactor model where underestimation of the number of factors is more problematic, the increase in α s produced greater accuracy. Again, as expected, for the single factor model where overestimation is a greater problem, accuracy decreased with an increase in α . Somewhat surprisingly, the decrease in accuracy for CDM was minimal. It is difficult to understand why CDM did not change more dramatically without having a greater understanding of the distributional properties of the RMSRs.

As demonstrated, changing α can increase or decrease accuracy depending on whether the parallel analysis methods tend to under- or overpredict the number of correct factors in a condition. We demonstrated the effect of a change in α on accuracy as a function of the generating model; however, we are not in a position to know the factor model when conducting a PA in practice. In addition, the effect of a change in α on accuracy would also be a function of sample size, values of factor loadings, and magnitude of factor correlations. Thus, it is impractical to change α depending on the characteristics of the factor analysis study. It could be argued that α for R-PA could be more liberal (perhaps, .10) for all analyses, or that α for CDM could be changed to balance Type I and Type II errors. This issue could be explored in future research. We suggest that it may be more beneficial to choose a conservative α (e.g., .05), but to introduce an effect size statistic to guide decision about the number of

factors. We discuss this alternative in the next section. We focus on R-PA in the next section because it produced overall greater accuracy in the current study; however, a similar argument for effect size statistics can be made for other PA methods.

Development of an Effect Size Statistic for PA. A problem with R-PA is that it essentially uses a sequential, hypothesis-testing strategy to make decisions about the number of factors. The process stops when the null hypothesis cannot be rejected. At this step, we conclude that the null hypothesis is correct, and no additional factors are required to reproduce the correlation matrix. The difficulty with this approach is that hypothesis testing is not designed to accept the null hypothesis. More specifically, the null hypothesis may not be rejected because of a lack of power. Accordingly, it would be helpful to have an effect size index that could indicate that a factor is strong in the sample even though we cannot conclude that this factor accounts for additional covariability among measures in the population using R-PA. In a related issue, an effect size statistic would also be useful to conclude that one or more of the factors detected by R-PA are relatively minor and may not be psychometrically important.

Comparing Methods Under Different Conditions. Our study examined a limited number of conditions as with any Monte Carlo study. It is potentially possible that the relative accuracies of methods would differ for other generating models (e.g., type of factor structure, factor loadings, and factor correlations) as well as sample size. It is important that PA methods are evaluated for conditions not explored in the current study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27*, 509-540.
- Crawford, A., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D. S., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70*, 885-901.
- 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.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The applications of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*, 291-314.

- 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.
- Green, S. B., Levy, R., Thompson, M. S., Lu, M., & Lo, W.-J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. *Educational and Psychological Measurement*, 72, 357-374.
- Green, S. B., Redell, N., Thompson, M. S., & Levy, R. (2016). Accuracy of revised and traditional parallel analyses for assessing dimensionality with binary data. *Educational and Psychological Measurement*, 76, 5-21.
- Green, S. B., Thompson, M. S., Levy, R., & Lo, W.-J. (2015). Type I and II error rates and overall accuracy of the revised parallel analysis method for determining the number of factors. *Educational and Psychological Measurement*, 75, 428-457.
- Harshman, R. A., & Reddon, J. R. (1983). Determining the number of factors by comparing real with random data: A serious flaw and some possible corrections. *Proceedings of the Classification Society of North America at Philadelphia*, 14-15.
- Hayashi, K., Bentler, P. M., & Yuan, K.-H. (2007). On the likelihood ratio test for the number of factors in exploratory factor analysis. *Structural Equation Modeling*, 14, 505-526.
- Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, 76, 537-549.
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13-43.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667-696.
- Reise, S. P., Moore, T. M., & Maydeu-Olivares, A. (2011). Targeted bifactor rotations and assessing the impact of model violations on the parameters of unidimensional and bifactor models. *Educational and Psychological Measurement*, 71, 684-711.
- Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, 23, 51-67.
- Ruscio, J., & Kaczetow, W. (2008). Simulating multivariate nonnormal data using an iterative algorithm. *Multivariate Behavioral Research*, 43, 355-381.
- Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, 24, 282-292.
- Socha, A. B., & Bandalos, D. L. (2015). *An investigation of the Hull, comparison data, and parallel analysis methods for determining the number of factors*. Unpublished manuscript.
- Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, 58, 541-568.