

Accuracy of Revised and Traditional Parallel Analyses for Assessing Dimensionality with Binary Data

Educational and Psychological
Measurement

2016, Vol. 76(1) 5–21

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164415581898

epm.sagepub.com



Samuel B. Green¹, Nickalus Redell¹,
Marilyn S. Thompson¹, and Roy Levy¹

Abstract

Parallel analysis (PA) is a useful empirical tool for assessing the number of factors in exploratory factor analysis. On conceptual and empirical grounds, we argue for a revision to PA that makes it more consistent with hypothesis testing. Using Monte Carlo methods, we evaluated the relative accuracy of the revised PA (R-PA) and traditional PA (T-PA) methods for factor analysis of tetrachoric correlations between items with binary responses. We manipulated five data generation factors: number of observations, type of factor model, factor loadings, correlation between factors, and distribution of thresholds. The R-PA method tended to be more accurate than T-PA, although not uniformly across conditions. R-PA tended to perform better relative to T-PA if the underlying model (a) was unidimensional but had some unique items, (b) had highly correlated factors, or (c) had a general factor as well as a group factor. In addition, R-PA tended to outperform T-PA if items had higher factor loadings and sample size was large. A major disadvantage of the T-PA method was that it frequently yielded inflated Type I error rates.

Keywords

factor analysis, parallel analysis, revised parallel analysis, binary data

Researchers use exploratory factor analysis (EFA) to find factors that can explain the covariation among measures in a parsimonious and meaningful way. Empirical criteria are frequently applied to suggest the number of factors that should be extracted;

¹Arizona State University, Tempe, AZ, USA

Corresponding Author:

Samuel B. Green, Arizona State University, PO Box 873701, Tempe, AZ 85287-3701, USA.

Email: samgreen@asu.edu

detailed overviews of these empirical strategies are available in the methodological literature (e.g., Crawford et al., 2010; Ruscio & Roche, 2012; Timmerman & Lorenzo-Seva, 2011; Velicer, Eaton, & Fava, 2000). Horn (1965) and others (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Preacher & MacCallum, 2003) have argued that parallel analysis (PA), possibly in conjunction with other criteria such as the scree test (Cattell, 1978), should be used to determine the number of factors.

Although there are a number of variations of PA, perhaps the most common approach involves the following steps: (a) conduct a principal component analysis (PCA) on sample data; (b) generate 100 or more comparison data sets with the same number of variables and sample size as the sample data, such that the variables are multivariate normally distributed in the population and uncorrelated; (c) perform a PCA on each of the comparison data sets; (d) calculate the mean eigenvalue for each sequential component extracted for these comparison data sets; and (e) determine the number of eigenvalues for the sample data that exceed the respective means of eigenvalues for the comparison data sets. This number is the estimated number of factors.

Green, Levy, Thompson, Lu, and Lo (2012) suggested a revised PA method (R-PA) to counter a previous criticism of traditional PA (T-PA) concerning the use of comparison data sets consisting of random normal data that are uncorrelated in the population (Harshman & Reddon, 1983; Turner, 1998). The R-PA method assesses the need for the k th factor by employing comparison data sets that are generated taking into account the existence of $k - 1$ factors. The k th eigenvalue for the sample data is compared with the results for k th eigenvalue for the comparison data sets. In the current study, we extend previous research evaluating R-PA with continuous data (Green et al., 2012; Green, Thompson, Levy, & Lo, 2014) to assessing R-PA for use with binary data.

Parallel Analysis Methods for Continuous Data

Psychometricians frequently have suggested two ways to improve the accuracy of PA. First, an extraction method based on the common factor model, such as principal axis factoring (PAF), is substituted for PCA in conducting a PA (Ford, MacCallum, & Tait, 1986; Mulaik, 2010). Some researchers have argued for the use of common factor analysis in PA because the underlying model allows for unreliability of measures, which is consistent with data collected in educational and psychological research (e.g., Fabrigar et al., 1999). Second, the eigenvalues for factors are compared with the 95th percentile of eigenvalues rather than the mean eigenvalue for random data sets (e.g., Buja & Eyuboglu, 1992; Glorfeld, 1995). The 95th percentile eigenvalue rule is a more stringent criterion and decreases the potential difficulty of overextraction of factors with PA (Zwick & Velicer, 1986).

Crawford et al. (2010) investigated the accuracy of PA with and without these recommended changes in a Monte Carlo study. No single PA approach was uniformly better than the others across conditions. Also, none of the methods were well

behaved; that is, their accuracies failed to consistently increase with increases in sample size, factor loadings, and number of variables per factor and with decreases in the correlations between factors.

One reason why these PA methods failed to behave well may be due to a problem described by Harshman and Reddon (1983) and Turner (1998). They argued that the use of reference distributions of eigenvalues based on data with no common factors (i.e., with uncorrelated variables) is only appropriate to reach conclusions about the relevance of the first factor. The proper reference distribution to reach a conclusion about the relevance of the k th factor should be based in general on data sets with $k - 1$ underlying factors.

Green et al. (2012) proposed R-PA that incorporates the use of an appropriately conditioned reference distribution of eigenvalues. With R-PA, the eigenvalue for the k th factor is compared with eigenvalues for data sets generated taking into account the existence of $k - 1$ factors. Ideally, the comparison data sets should be generated based on the population loadings of these $k - 1$ factors. Because the population factor loadings are unknown, sample factor loadings are substituted for the population values in conducting this revision to PA. Green et al. (2012) conducted a Monte Carlo study to compare the accuracy of T-PA and R-PA methods using either PCA or PAF in conjunction with either the mean or 95th percentile eigenvalue rule. R-PA using PAF and the 95th percentile rule had relatively high accuracy and behaved better statistically than the other methods. T-PA using PAF and the 95th percentile rule also demonstrated relatively high accuracy, but was not quite as well behaved.

Green et al. (2014) considered PA using PAF and the 95th percentile eigenvalue rule as a series of hypotheses tests. They argued that within a hypothesis testing framework, this traditional PA approach employs the wrong sampling distribution. In comparison, the *revised PA* method applying PAF and the 95th percentile eigenvalue rule was described as involving tests of null hypotheses that the data have no more than $k - 1$ underlying factors. At any step, the null hypothesis of no more than $k - 1$ underlying factors is rejected at the .05 level if the k th eigenvalue for the sample data is positive and exceeds the 95th percentile of eigenvalues for the k th factor of the comparative data sets. In this sequential process, k initially equals 1. If the hypothesis is rejected, k is increased by 1 to assess the null hypothesis that the variables have no more than 1 underlying factor in the population. If this hypothesis is rejected, k is again increased by 1, and the process is continued, evaluating at each step the null hypothesis of no more than $k - 1$ factors, until the null hypothesis is not rejected. At this point, the researcher may conclude that the number of factors is equal to the k of the previous step.

Green et al. (2014) conducted a Monte Carlo study to assess the accuracy of R-PA and T-PA (using PAF and the 95th percentile eigenvalue rule) as well as traditional likelihood ratio tests (LRTs; Hayahi, Bentler, & Yuan, 2007). Overall, the PA approaches tended to outperform the LRT methods. T-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, and conditions in which some of the variables were not a function of any common factors.

The authors further investigated the PA methods by examining their empirical Type I error rates, relative to the nominal alpha of .05, and their empirical powers. These results can be summarized as follows: (a) Empirical alphas with the T-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. (b) The empirical alphas with R-PA were in most conditions below .05, with only one exceeding the value of .064 (i.e., .077). The empirical alphas tended to be overly conservative with high factor loadings and more than a single factor. (c) In conditions with lower factor loadings, T-PA had greater empirical power than the revised method, which explains its greater overall accuracy in these conditions. However, this greater power was due to inflated alphas. (d) R-PA showed greater accuracy under a variety of conditions. In these conditions, the greater accuracy was due to greater power rather than inflated alphas.

Independent of and concurrent with the work of Green and his colleagues (Green et al., 2012; Green et al., 2014), Ruscio and Roche (2012) proposed and evaluated an alternative approach for generating comparison data sets based on $k-1$ factors. Their method takes a more nonstandard approach than the one proposed by Green and his colleagues (Green et al., 2012; Green et al., 2014) and involves computing root mean squared differences between the eigenvalues for the sample data and the comparison data set and conducting a series of Mann–Whitney U tests with a liberal alpha of .30. They conducted a Monte Carlo study and found that their method performed well across a range of conditions.

Parallel Analysis for Ordered Categorical Data

PA can be applied to assess the number of factors underlying items on a test that have binary scales or ordered categorical response scales (e.g., Likert scale). The nature of the research on PA for ordered categorical data appears to differ depending on why it is being conducted. In some studies, PA is an intermediate step in EFA to determine the number of factors to extract. In other studies, PA is viewed as a method to assess whether a single dimension underlies a set of items, and, if so, then unidimensional item response theory can be applied.

Traditional Parallel Analysis and the Choice of Correlations to Be Analyzed

Pearson product–moment correlations can underestimate the relationships between the ordered categorical variables (e.g., Bollen & Barb, 1981). Underestimation is likely to be greater to the extent that the number of response categories is limited (e.g., binary) and the distributions of the categorical variables differ between variables (e.g., items are highly skewed in opposite directions). Thus, the pattern of Pearson product–moment correlations in a matrix of ordered categorical items can be

influenced by the disparity between the distributions of these items, particularly if they have binary responses. The implication is that spurious factors can occur that represent differences in distribution of variables rather than the true dimensionality of items (Flora & Curran, 2004; Gorsuch, 1983). For binary items, these spurious factors are frequently referred to as difficulty factors.

As an alternative, polychoric correlations can be computed between ordered categorical items; tetrachoric correlations are a special case of polychoric correlations in which items have binary responses. In computing polychoric correlations, it is assumed the ordered categorical items have underlying variables that are normally distributed. Polychoric correlations are estimates of the correlations between the latent, normally distributed variables.

A number of studies have been conducted to assess the accuracy of traditional PA with ordered categorical items (e.g., Cho, Li, & Bandalos, 2009; Garrido, Abad, & Ponsoda, 2012; Green, 1983; Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Weng & Cheng, 2005). These studies varied in the conditions explored in generating the ordered categorical data and in the methods used in conducting PA. For example, some researchers (Weng & Cheng, 2005; Tran & Forman, 2009) assessed the accuracy of PA to evaluate unidimensionality (i.e., one common factor), given its importance in item response theory, and others (Cho et al., 2009; Garrido et al., 2012; Green, 1983; Timmerman & Lorenzo-Seva, 2011) assessed the accuracy of PA to evaluate dimensionality in general.

A focus of these studies (except for Green, 1983) was on the relative accuracy of PA when the analyses were conducted with Pearson product–moment correlations versus polychoric correlations (or tetrachoric correlations for binary responses). The findings and recommendations of these studies varied. For example, Tran and Forman (2009) concluded that PA yielded unsatisfactorily low accuracies with either type of correlation coefficients. Cho et al. (2009) and Weng and Cheng (2005) showed that under most conditions, PA with product–moment correlations yielded greater accuracy. However, the most recent and extensive studies by Timmerman and Lorenzo-Seva (2011) and Garrido et al. (2012) concluded that PA with polychoric correlations yielded better results, although problems such as failure to converge and nonpositive definite matrices can occur in the estimation of polychoric correlation matrices.

Parallel Analysis for Binary Items Using Comparative Data Sets with Dimensionality

All of the cited studies involving ordered categorical items focused on the accuracy of traditional PA. Interestingly, Drasgow and Lissak (1983) examined more than 30 years ago the effectiveness of PA to assess unidimensionality using a comparative data set that was structured similarly to ones used with the R-PA approach. With the Drasgow–Lissak method, a comparative item data set is computer generated based on item parameters estimated from the sample data. Correlation matrices are computed

for the binary data from the sample data set and the comparative data set and factor analyzed. The eigenvalues from these two factor analyses are displayed on a scree plot and visually examined to make a decision about whether one factor or multiple factors underlie the sample data set. Presumably because this method was proposed prior to the advent of high speed computers, only a single comparative data set is generated. The limited empirical results presented in their article suggested the approach had potential.

Budescu, Cohen, and Ben-Simon (1997) and Finch and Monahan (2008) presented research that extended the work of Drasgow and Lissak (1983). Both sets of authors recognized the insufficiency of generating a single comparative data set and sought to remedy this problem (as well as other potential difficulties). The method by Budescu et al. (1997) involves generating an expected matrix of correlations assuming that a three-parameter logistic model underlies the sample data and focuses on eliminating items that make the test not unidimensional. Finch and Monahan (2008) revised the Drasgow–Lissak method by introducing a bootstrap method to assess dimensionality. Both Budescu et al. (1997) and Finch and Monahan (2008) conducted Monte Carlo studies to assess their revisions to the Drasgow–Lissak method, but narrowed their studies to the accuracy of identifying whether a single factor underlies the sample data.

Objectives of this Study

Many studies have investigated the accuracy of PA to evaluate the dimensionality of a set of variables and generally found it to be one of the best methods for evaluating the number of underlying factors (e.g., see summaries by Fabrigar et al., 1999; Preacher & MacCallum, 2003). Fewer studies have examined PA with ordered categorical items. Overall, these studies suggest that PA can be an effective method and PA is likely to yield better results with polychoric correlations (Garrido et al., 2012; Timmerman & Lorenzo-Seva, 2011). Although not explicitly investigated, there seems to be some belief that PA may not be as effective with binary items (see Tran & Formann, 2009).

Almost all the research conducted using ordered categorical data has involved assessing the effectiveness of traditional PA. The exception is the research by Drasgow and Lissak (1983), who presented a PA method that generated a comparative data set with a single underlying dimension rather than uncorrelated variables as with traditional PA. However, their method was limited to the assessment of unidimensionality and was evaluated using a very limited Monte Carlo investigation. Budescu et al. (1997) and Finch and Monahan (2008) recommended modifications to the Drasgow–Lissak method, but limited their Monte Carlo investigation of their methods to the assessment of unidimensionality and suggested revisions that take PA in a very different direction (e.g., elimination of items not conforming to unidimensionality).

The purpose of our research is to assess the accuracy of T-PA and R-PA in estimating the number of factors underlying binary data. The research extends the work of Drasgow and Lissak (1983) and Green and his colleagues (Green et al., 2012;

Green et al., 2014) by exploring a revised PA method to assess a variety of multifactor models as well as unidimensional ones. In addition, we consider Type I and II errors in the stepwise PA process to diagnose problems in the methods, similar to Green et al. (2014) and Finch and Monahan (2008). Finally, we investigate the effectiveness of PA methods for binary data in comparison with continuous data by comparing the results of the current study with those of a similarly structured previous study (Green et al., 2014).

Method

Design

In all conditions, the number of items was eight, and the number of response categories for all items was two. Five data generation factors were manipulated to produce 52 conditions. The five generation factors were as follows:

- *Number of observations* (N_O): The number of observations was set at 200 or 400.
- *Type of factor model*: Data were generated based on five types of factor models: (a) a zero-factor model in which all items were a function of only error; (b) a unidimensional model in which all items loaded on a single factor (referred to as one-factor model for all items); (c) a unidimensional model in which half of the items loaded on a single factor, and the other half were a function of only error (one-factor model with unique items); (d) a two-factor, perfect-clusters model (Browne, 2001), with half of the items loading on one factor and the other half of the items loading on the second factor; and (e) a two-factor, bifactor model, with all items loading on a general factor and half of the items also loading on a group factor.
- *Factor loadings* (λ): For a one-factor model for all items, loadings for items were either .5s or .7s. For a one-factor model with unique items, loadings for items on the single factor were either .5s or .7s. For a two-factor, perfect-clusters model, the nonzero loadings on the two factors were either all .5s or all .7s. For a two-factor, bifactor model, the eight items had either all .5s or all .7s on the general factor, and four of the items had .5s on the group factor (i.e., loadings were not varied across conditions on the group factor).
- *Factor correlations* (ρ_{F1F2}): For a two-factor, perfect-clusters model, the correlation between factors was 0, .5, or .8. For a two-factor, bifactor model, the correlation between factors was always 0.
- *Thresholds* (τ) *distributions*: The thresholds were either uniform or mixed across items. Uniform thresholds had 0s across all items, whereas mixed thresholds had $-.5$ on four items and $+.5$ on the remaining four items. For a single-factor with unique items, the $-.5$ and $+.5$ thresholds were split evenly across the four items with nonzero loadings on the factor as well as across the four items with zero loadings on the factor. For a two-factor, perfect-clusters

model, the $-.5$ and $+.5$ thresholds were split evenly across the four items with nonzero loadings for each factor. For a two-factor, bifactor model, the $-.5$ and $+.5$ thresholds were split evenly across the four items on the group factor and across the remaining four items.

Data Generation and Analyses

Data were generated and analyzed using SAS 9.2. We used RANNOR to generate normally distributed data, PROC FREQ to compute tetrachoric correlations, and PROC FACTOR to conduct factor analyses.

Sample Data Sets. For each combination of manipulated conditions, 1,000 sample data sets were generated with a common factor model. The factors and the errors in the model were generated to be normally distributed. Thresholds were then imposed on the continuous item data to yield sample data sets of binary item scores. Tetrachoric correlation matrices were computed for each sample data set, and these correlations were analyzed using PAF. If correlation matrices were positive definite and the factor solution yielded no out-of-bound estimates, eigenvalues were retained to be compared with those for the comparison data sets.

Comparison Data Sets. For traditional PA, 100 comparison data sets were generated for each of the 1,000 sample data sets. The variables within data sets were generated to be normally distributed and uncorrelated using RANNOR. These 100 data sets had the same number of observations as the sample data sets. Thresholds were imposed to create binary scores for the comparison data sets that were consistent with the binary score distributions of the sample data sets. Tetrachoric correlation matrices were computed and analyzed using PAF with multiple R^2 s along the diagonal.

For revised PA, the comparison data sets for the null hypothesis of zero factors were the comparison data sets for traditional PA. The comparison data sets for the null hypothesis of $k-1$ or fewer factors were generated based on loadings of the $k-1$ factors from the factor analyses of the sample data sets. The comparison data sets had the same number of observations as the sample data sets. Thresholds were imposed to create binary scores for the comparison data sets that were consistent with the binary score distributions of the sample data sets.

If a correlation matrix was not positive definite or a factor analysis yielded an out-of-bound estimate for a comparison data sets for a PA method, that comparison data sets was excluded. If a comparison data sets was eliminated for one PA method (e.g., T-PA), a comparison data sets also was deleted for the other PA method (e.g., R-PA) in order to hold constant the number of comparison data sets across the two PA methods. The 95th percentiles of eigenvalues of the factors for each sets of comparison data sets were retained for the traditional and revised PA methods. The eigenvalues for the sample data sets were then compared with the 95th percentiles of eigenvalues

for the comparative data sets to yield an estimated number of factors for traditional and revised PA methods.

Criteria. For each condition, we computed the proportion of sample data sets in which T-PA and R-PA accurately estimated the number of factors, underestimated the number of factors, and overestimated the number of factors. If we view the two PA methods as a series of hypothesis tests, the accuracy of these methods should not exceed .95. An accuracy of 95% occurs for a model with k underlying factors if (a) the power for tests of null hypotheses of fewer than k factors have powers approaching 1.0 and (b) the alpha for the test of the null hypothesis of k factors is at the nominal level of .05. The nominal alpha is at the .05 level because the sample eigenvalues were compared with the 95th percentile of eigenvalues for the comparative data sets. Implicit within this framework, a method has inflated alphas if the percent of data sets that overestimate the number of factors is greater than .05. Similarly, underestimation of the number of factors may be viewed as a lack of power. However, in comparing methods, we must be careful in reaching conclusions about the relative power of the two PA methods in that perceived greater power could be due to inflated alphas for a method.

Results

The proportion of improper solutions for sample data sets or comparison data sets exceeded 4% for only one condition (34.5% improper solutions for the condition with $N_O = 200$, bifactor model with $\lambda = .7$ on the general factor, and heterogeneous τ). In general, the conditions with two underlying factors, factor loadings of .7, and heterogeneous thresholds produced higher percentages of improper solutions. Given the small number of improper solutions across the large majority of conditions, we concluded that the deletion of data sets having analyses with improper solutions had a minimal effect on our results.

Accuracies

We present accuracies (i.e., proportions of sample data sets with correctly identified number of factors) for the various conditions in Tables 1 and 2. For conditions with no underlying factors, T-PA and R-PA must have the same accuracies. These results offer some validation of the Monte Carlo program in that accuracies ranged from .941 to .959, as one would expect given the choice of the 95th percentile rule for eigenvalues of comparison data sets.

For conditions with one-factor models for all items, the accuracies for R-PA and T-PA were very high, exceeding .910 for all conditions. It is interesting to examine these results in greater detail. None of the PA methods ever underestimated the number of factors. In other words, the empirical powers to reject the null hypothesis of zero factors were all 1.0. In contrast, the two PA methods differed in terms of the proportion of data sets in which they yielded overestimates of the number of factors.

Table 1. Proportion of Sample Data Sets in which T-PA and R-PA Underestimated, Correctly Estimated, and Overestimated the Number of Factors for Generation Models With Zero or One Factor.

Thresholds	λ	$N_0 = 200$						$N_0 = 400$					
		T-PA			R-PA			T-PA			R-PA		
		Hits	Under	Over	Hits	Under	Over	Hits	Under	Over	Hits	Under	Over
Same	—	.947	—	.053	.947	—	.053	.941	—	.059	.941	—	.059
		.959	—	.041	.959	—	.041	.954	—	.046	.954	—	.046
Same	.5	.930	.000	.070	.952	.000	.048	.967	.000	.033	.950	.000	.050
		.993	.000	.007	.962	.000	.038	.998	.000	.002	.959	.000	.041
Different	.5	.914	.000	.086	.954	.000	.046	.956	.000	.044	.961	.000	.039
		.968	.000	.032	.966	.000	.034	.993	.000	.007	.950	.000	.050
Same	.5	.764	.075	.161	.866	.075	.059	.845	.000	.155	.952	.000	.048
		.832	.000	.168	.954	.000	.046	.870	.000	.130	.949	.000	.051
Different	.5	.684	.161	.155	.791	.161	.048	.819	.005	.176	.935	.005	.060
		.777	.000	.223	.942	.000	.058	.805	.000	.195	.950	.000	.050

Note. R-PA = revised parallel analysis; T-PA = traditional parallel analysis.

Zero-factor model

One-factor model for all items

One-factor model with unique items

Table 2. Proportion of Sample Data Sets in which T-PA and R-PA Underestimated, Correctly Estimated, and Overestimated the Number of Factors for Generation Models With Two Factors.

Thresholds	λ	$N_O = 200$						$N_O = 400$					
		T-PA			R-PA			T-PA			R-PA		
		Hits	Under	Over	Hits	Under	Over	Hits	Under	Over	Hits	Under	Over
Same	.5	.822	.042	.136	.862	.091	.047	.949	.001	.050	.955	.001	.044
Same	.7	.961	.000	.039	.962	.000	.038	.992	.000	.008	.954	.000	.046
Different	.5	.776	.082	.142	.788	.179	.033	.917	.000	.083	.949	.002	.049
Different	.7	.863	.000	.137	.976	.000	.024	.968	.000	.032	.955	.000	.045
		Two-factor, perfect-clusters model with $\rho_{F1F2} = 0$											
Same	.5	.496	.421	.083	.432	.539	.029	.808	.140	.052	.800	.161	.039
Same	.7	.951	.010	.039	.952	.008	.040	.994	.000	.006	.958	.000	.042
Different	.5	.392	.508	.100	.276	.702	.022	.707	.224	.069	.683	.287	.030
Different	.7	.873	.029	.098	.940	.034	.026	.968	.000	.032	.967	.000	.033
		Two-factor, perfect-clusters model with $\rho_{F1F2} = .5$											
Same	.5	.098	.880	.022	.079	.915	.006	.148	.843	.009	.158	.833	.009
Same	.7	.267	.720	.013	.421	.558	.021	.573	.417	.010	.809	.153	.038
Different	.5	.122	.836	.042	.060	.932	.008	.126	.859	.015	.129	.866	.005
Different	.7	.253	.704	.043	.287	.701	.012	.510	.463	.027	.653	.307	.040
		Two-factor, bifactor model											
Same	.5	.291	.676	.033	.294	.687	.019	.587	.396	.017	.624	.349	.027
Same	.7	.373	.672	.010	.611	.364	.025	.763	.233	.004	.901	.065	.034
Different	.5	.250	.683	.007	.199	.785	.016	.525 ^a	.438	.037	.522 ^a	.454	.024
Different	.7	.284	.634	.082	.295	.701	.004	.699	.268	.033	.797	.169	.034

Note. R-PA = revised parallel analysis; T-PA = traditional parallel analysis.

^aWe ran one extra condition to assess whether R-PA would outperform T-PA if sample size was further increased. With a sample size of 600, the accuracies, proportion underpredicted, and proportion overpredicted were .691, .278, and .031, respectively, for T-PA and .725, .230, and .045, respectively, for R-PA.

Because there was no underestimation of the number of factors, these proportions are equivalent to empirical alphas for testing the null hypothesis of one or fewer factors. The empirical alphas for R-PA were relatively close to the nominal alphas of .05, ranging in value from .034 to .050, and translate into accuracies ranging in value from .950 to .966 (i.e., $1 - \alpha$). On the other hand, the empirical alphas for T-PA ranged in value from .002 to .086, and, thus, the accuracies were from .914 to .998 (i.e., $1 - \alpha$). The alphas for T-PA tended to be greater than .05 with a sample size of 200 and factor loadings of .5s. In contrast, the alphas were negatively biased when loadings were .7s. Thus, T-PA was more accurate than R-PA for models with loadings of .7s, but from a hypothesis testing perspective, the greater accuracy was due to the overly conservative Type I error rates.

For all eight conditions investigating one-factor models with unique items, R-PA substantially outperformed T-PA. The average difference in accuracies was .118. The proportions of data sets with underestimated number of factors for these conditions were identical for R-PA and T-PA. Consequently, the reason T-PA had poor accuracy was due to overestimation of the number of factors. Across all 8 conditions, the proportion of overestimation for T-PA was between .130 and .223. Given an alpha of .05, the proportion of overestimation should not exceed .05. In contrast to those from T-PA, the proportions of overestimation for R-PA ranged from .046 to .060.

The results of conditions with two-factor, perfect-clusters models differed depending on the magnitude of the correlation between factors. When the factor correlation was 0, R-PA and T-PA performed similarly (within .02 of each other) in four of the eight conditions. R-PA performed better in three of the remaining four conditions. These differences are directly tied to Type I error rates. The proportion of data sets with an overestimated number of factors should not exceed .05 given the sets alpha of .05. However, in the three conditions in which T-PA performed relatively poorly, the proportions of data sets with an overestimated number of factors were .083, .136, and .137 for T-PA versus .049, .047, and .024 for R-PA. In the one condition in which R-PA performed relatively poorly, T-PA had higher accuracy because of its overly conservative Type I error rate of .008 (vs. .046 for R-PA).

When the factor correlation was .5 for two-factor, perfect-clusters models, the results are less clear. T-PA performed better than R-PA in four conditions; R-PA yielded more accurate results in one condition; and the two methods yielded approximately the same degree of accuracy for the remaining three cases (within .02 of each other). The proportions of data sets with an underestimated number of factors were either similar or greater for R-PA across these eight conditions. These differences might be attributed to greater power of T-PA to reject the null hypothesis of one or fewer factors; however, based on the previous results, it is quite possible that these differences are due to inflated alphas. In contrast, the proportions of data sets with an overestimated number of factors were .006 to .100 for T-PA and .022 to .042 for R-PA. These results indicated that T-PA had inflated empirical alphas for the null hypothesis of two or fewer factors in three conditions and deflated empirical alphas for two other conditions (in which powers for earlier tests in the sequence were 1.0).

Both methods were less accurate in conditions with perfect-clusters models and a correlation of .8 between factors than those in previous conditions. T-PA performed better than R-PA in one condition; R-PA yielded more accurate results in four conditions; and the two methods yielded approximately the same degree of accuracy in the remaining three conditions (within .02 of each other). In all four conditions in which R-PA outperformed T-PA, the factor loadings were .7s. The differences in accuracies were substantial (.143 to .236) in three of four of these conditions and were due to the lower proportion of data sets with an underestimated number of factors.

Finally, for the eight conditions with two-factor, bifactor models, the relative accuracies for T-PA and R-PA were generally similar to those for conditions with highly correlated, perfect-clusters models. The most substantial differences were for three conditions in which the factor loadings were .7s on the general factor; these differences were in favor of R-PA.

Regardless of which PA method was applied, accuracies tended to deteriorate with heterogeneous thresholds. We suspect the deterioration is due to sampling variability in the eigenvalues.

Comparison of Accuracies for Parallel Analysis on Binary and Continuous Data

We compared the accuracies for our current study involving binary data with the accuracies from a previous study based on continuous, normal data (Green et al., 2014). We present these results in Table 3. It should be noted that the factor loadings and correlations between factors were for the underlying normal scores of the binary data. In most cases, the performance with binary data suffered relative to continuous normal data, which is consistent with the notion that binary or ordered discrete data pose challenges to factor modeling in general (e.g., Mislavy, 1986; Olsson, 1979). Overall, it is quite apparent that we must be careful in designing studies with binary data relative to continuous data. To yield results more similar to those obtained with normally distributed data, we need to design studies with measures that have a higher saturation of the underlying factor(s) and/or with larger sample sizes.

Conclusion

In general, the results are supportive of the revised PA method, although T-PA can yield more accurate results under some conditions. The advantages of R-PA are threefold. First, R-PA was more accurate than T-PA in most conditions, particularly for conditions most likely observed with well-designed studies. By well-designed studies, we mean those with measures with high factor loadings and large sample sizes. Second, R-PA demonstrated better control of Type I error rates in the current study with binary data as well as in a previous study with continuous, normal data (Green et al., 2014). Third, R-PA is more defensible on theoretical grounds. To assess

Table 3. Accuracies for Traditional and Revised Parallel Analyses for Binary Data From the Current Study and for Continuous Data based on a Previous Study (Green et al., 2014).

λ	$N_O = 200$						$N_O = 400$					
	T-PA			R-PA			T-PA			R-PA		
	Binary		Normal	Binary		Normal	Binary		Normal	Binary		Normal
	$=\tau$	$\neq\tau$		$=\tau$	$\neq\tau$		$=\tau$	$\neq\tau$		$=\tau$	$\neq\tau$	
.5	.930	.914	.997	.952	.954	.969	.967	.956	1.000	.950	.961	.959
.7	.993	.968	1.000	.962	.966	.972	.998	.993	1.000	.959	.950	.958
.5	.764	.684	.850	.866	.791	.936	.845	.819	.879	.952	.935	.939
.7	.832	.777	.897	.954	.942	.936	.870	.805	.911	.949	.950	.936
.5	.822	.776	.984	.862	.788	.983	.949	.917	1.000	.955	.949	.997
.7	.961	.863	1.000	.962	.976	.998	.992	.968	1.000	.954	.955	1.000
.5	.496	.392	.876	.432	.276	.899	.808	.707	.997	.800	.683	.984
.7	.951	.873	1.000	.952	.940	.990	.994	.968	1.000	.958	.967	.998
.5	.098	.122	.100	.079	.060	.204	.148	.126	.187	.158	.129	.481
.7	.267	.253	.578	.421	.287	.958	.573	.510	.967	.809	.653	.988
.5	.291	.250	.610	.294	.199	.820	.587	.525	.935	.624	.522	.988
.7	.373	.284	.820	.611	.295	.982	.763	.699	.997	.901	.797	.996

Note. R-PA = revised parallel analysis; T-PA = traditional parallel analysis.

a hypothesis about the adequacy of $k - 1$ factors, it is necessary to create a sampling distribution of eigenvalues based on an underlying structure of $k - 1$ factors.

The results indicated that R-PA performed best relative to T-PA for one-factor models with unique indicators and for highly correlated two-factor models and bifactor models with high factor loadings and large sample sizes. In contrast, the results suggest that the T-PA can yield superior results that are not due to inflated Type I error rates if factor loadings are small and sample sizes are small. With small sample sizes, the factor loadings used to conduct R-PA are less stable, producing greater variability in the eigenvalue distributions based on the comparative data sets, which potentially reduces the accuracy of R-PA. In addition, the advantage of R-PA over T-PA diminishes as the population factor loadings approach zero in that T-PA essentially assumes that these loadings are zero in contrast to R-PA, which uses estimates of these loadings. It is possible that R-PA may yield improved results if the number of comparison data sets is increased beyond the typical number of 100 for conditions with small samples and/or low factor loadings.

Additional research is required to evaluate the revised PA method under additional data conditions likely to occur in practice. In particular, other factor structures, such as those with varied factor loadings, should be investigated. Also, it is important to assess whether the findings with binary data extend to ordered categorical data with more than two response options. In addition, the performance of revised PA should be explored for conditions in which measures have nonnormal, continuous distributions or in which item data have nonnormal, continuous distributions underlying ordered categorical responses.

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

- Bollen, K. A., & Barb, K. H. (1981). *Pearson's r and coarsely categorized measures*. *American Sociological Review*, *46*, 232-239.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111-150.
- Budescu, D. V., Cohen, Y., & Ben-Simon, A. (1997). A revised modified parallel analysis for the construction of unidimensional item pools. *Applied Psychological Measurement*, *21*, 233-252.
- Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, *27*, 509-540.

- Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Plenum.
- Cho, S.-J., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement, 69*, 748-759.
- 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.
- Dragow, F., & Lissak, R. I. (1983). Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied Psychology, 68*, 363-373.
- 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.
- Finch, H., & Monahan, P. (2008). A bootstrap generalization of modified parallel analysis for IRT dimensionality assessment. *Applied Measurement in Education, 21*, 119-140.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466-491.
- 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.
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2012). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods, 18*, 454-474.
- 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.
- Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Green, S. B. (1983). Identifiability of spurious factors using linear factor analysis with binary items. *Applied Psychological Measurement, 7*, 139-147.
- 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., Thompson, M. S., Levy, R., & Lo, W.-J. (2014). 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*. Advance online publication. doi:10.1177/0013164414546566
- 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.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.
- Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics, 11*, 3-31.
- Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC Press.

- Olsson, U. (1979). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research, 14*, 485-500.
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*, 13-43.
- 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.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*, 209-220.
- Tran, U. S., & Formann, A. K. (2009). Performance of parallel analysis in retrieving unidimensionality in the presence of binary data. *Educational and Psychological Measurement, 69*, 50-61.
- 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.
- 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 Jackson at seventy* (pp. 41-71). Norwell, MA: Kluwer Academic.
- Weng, L.-J., & Cheng, C.-P. (2005). Parallel analysis procedure with unidimensional binary data. *Educational and Psychological Measurement, 65*, 697-716.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.