Parallel Analysis with Unidimensional Binary Data

Li-Jen Weng and Chung-Ping Cheng

Educational and Psychological Measurement 2005; 65; 697
DOI: 10.1177/0013164404273941

PARALLEL ANALYSIS WITH UNIDIMENSIONAL BINARY DATA

LI-JEN WENG
CHUNG-PING CHENG
National Taiwan University

The present simulation investigated the performance of parallel analysis for unidimensional binary data. Single-factor models with 8 and 20 indicators were examined, and sample size (50, 100, 200, 500, and 1,000), factor loading (.45, .70, and .90), response ratio on two categories (50/50, 60/40, 70/30, 80/20, and 90/10), and types of correlation coefficients (phi and tetrachoric correlations) were manipulated. The results indicated that parallel analysis performed well in identifying the number of factors. The performance improved as factor loading and sample size increased and as the percentages of responses on two categories became close. Using the 95th and 99th percentiles of the random data eigenvalues as the criteria for comparison in parallel analysis yielded higher correct rate than using mean eigenvalues.

Keywords: *factor analysis; parallel analysis; dichotomous variables; number of factors; unidimensionality*

Determining the number of factors plays a critical role in exploratory factor analysis. The over- or underextraction of common factors can result in inappropriate conclusions and even affect the development of psychological theories (Comrey & Lee, 1992; Gorsuch, 1983; Harman, 1976). The present research focused on the application of parallel analysis to unidimensional binary data. Because other methods of deciding the number of factors have been shown to be unsatisfactory for binary data (e.g., Bernstein & Teng,

Chung-Ping Cheng is now in the Department of Psychology at Kaohsiung Medical University, Kaohsiung, Taiwan. This research was supported in part by Grant NSC 91-2413-H-002-015 from the National Science Council of Taiwan. We wish to thank the editor and two referees for helpful comments and suggestions on an earlier draft. Correspondence concerning this article should be addressed to Li-Jen Weng, Department of Psychology, National Taiwan University, Taipei 106, Taiwan, Republic of China; e-mail: ljweng@ntu.edu.tw.

Educational and Psychological Measurement, Vol. 65 No. 5, October 2005 697-716
DOI: 10.1177/0013164404273941

1989), we hoped that the well-received method of parallel analysis could be a solution to the problem.

Researchers have proposed different methods for detecting the number of factors since Spearman's introduction of factor analysis. For example, the eigenvalue > 1 rule (also known as the Kaiser-Guttman rule) and the scree test are among the methods commonly used (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Ford, MacCallum, & Tait, 1986; Wang & Weng, 2002; Weng, 1995). Recent research has found that the parallel analysis proposed by Horn (1965) performed well (e.g., Reilly & Eaves, 2000; Sarff, 1997; Velicer, Eaton, & Fava, 2000; Wang, 2001; Zwick & Velicer, 1986). Because popular statistical software packages such as SAS and SPSS have not included parallel analysis as an option for factor analysis, researchers have developed programs to facilitate the implementation of the method (Enzmann, 1997; Kaufman & Dunlap, 2000; Lautenschlager, 1989b; Longman, Cota, Holden, & Fekken, 1989a; O'Connor, 2000; Thompson & Daniel, 1996). The increasing research effort suggests that parallel analysis has been well received as a method for determining the number of common factors.

Horn (1965) proposed the method of parallel analysis to adjust for the frequently used decision rule for the number of factors of eigenvalue > 1. The eigenvalue > 1 method assumes the correlation matrix analyzed to be the population correlation matrix. Horn argued that the effects of sampling errors on eigenvalues of sample correlation matrices should be taken into account in applying the eigenvalue > 1 rule in determining the number of factors. Horn therefore suggested comparing the eigenvalues of the sample correlation matrix with those obtained from the random data correlation matrix of the same number of variables and sample size as the criterion for deciding the number of factors. The mean eigenvalues from several random data correlation matrices were recommended to reflect the effects of sampling errors. The number of factors equals the number of sample eigenvalues greater than the mean eigenvalues of random data correlation matrices.

Following Horn's (1965) proposition, Humphreys and Ilgen (1969) and Humphreys and Montanelli (1975) found that the new method performed well in determining the number of factors. To facilitate the implementation of parallel analysis, regression equations and interpolation tables for the estimation of mean eigenvalues have been developed (Allen & Hubbard, 1986; Keeling, 2000; Lautenschlager, 1989a; Lautenschlager, Lance, & Flaherty, 1989; Longman, Cota, Holden, & Fekken, 1989b; Montanelli & Humphreys, 1976). In addition to the mean eigenvalues, the 95th percentile eigenvalues from random data were proposed as another criterion in parallel analysis, and regression estimation and the interpolation method were suggested accordingly (Cota, Longman, Holden, Fekken, & Xinaris, 1993; Longman et al., 1989b).

Glorfeld (1995) noticed from Zwick and Velicer's (1986) simulation that parallel analysis had a slight tendency to overextract one or two factors. He suggested generating a very large number of random correlation matrices to fully represent the complete sampling distribution of eigenvalues. A desired statistical significance level could then be chosen as the criterion to adjust for the tendency of overextraction with Horn's (1965) parallel analysis. The 95th percentile of eigenvalues from random correlation matrices has been proposed as a substitute of average eigenvalues to avoid the tendency of overextraction (Buja & Eyuboglu, 1992; Glorfeld, 1995). Wang's (2001) simulation study indicated that except for the regression equations suggested by Allen and Hubbard (1986) and Montanelli and Humphreys (1976), parallel analysis was able to recover the correct number of common factors with Likert-type scales whether mean eigenvalues or the 95th-percentile eigenvalues were used as the basis for comparison.

Factor analysis is a frequently used statistical method in scale development and validity studies. Among the response formats adopted by researchers, binary response format is commonly used, especially in cognitive or psychotic symptom measures. Parry and McArdle (1991) simulated data on the basis of a single-factor model with eight dichotomous indicators and studied the performance of factor analysis on unidimensional binary data. In their well-designed investigation, they examined the behaviors of four methods for estimating factor loadings with phi and tetrachoric correlations, but they did not explore the issue of determining the number of factors with binary data. Collins, Cliff, McCormick, and Zatkin (1986) applied a modification of the scree test to inspect the dimensionality of items constructed by two-parameter logistic modeling. The method tended to extract too many factors with either phi correlations or tetrachoric correlations. Roznowski, Tucker, and Humphreys (1991) compared the performance of three methods for determining the number of factors on binary variables but did not include parallel analysis.

Green (1983) used parallel analysis and the scree test to discover the dimensionality of items constructed by three-parameter logistic modeling. Parallel analysis was found to extract too many factors when phi correlations were used in analysis. Green's parallel analysis used the regression equation suggested by Montanelli and Humphreys (1976) for estimating mean eigenvalues of random data correlation matrices. However, this equation was found to perform poorly in identifying the number of factors (Wang, 2001).

Bernstein and Teng (1989) scrutinized the performance of two methods of deciding the number of factors on a single-factor model with 20 variables measured on dichotomous scales and 4-point scales. The eigenvalue > 1 rule and the maximum likelihood chi-square test were found to overestimate the number of factors. Bernstein and Teng therefore concluded that methods for determining the number of factors developed for continuous variables were inappropriate for noncontinuous data. However, Wang's (2001) simulation study suggested that parallel analysis worked well for Likert-type rating scales. Whether the mean eigenvalues or the 95th-percentile eigenvalues were used as the baseline for comparison, parallel analysis was able to suggest the correct number of factors, except when the regression equations proposed by Allen and Hubbard (1986) and Montanelli and Humphreys (1976) were used to estimate the mean eigenvalues of random data correlation matrices.

Turner (1998) evaluated the performance of parallel analysis against continuous data and 4-point, Likert-type data and suggested that the most useful application of parallel analysis perhaps was to determine if a scale or a subscale was unidimensional. In view of the performance of parallel analysis in determining the number of factors over the past research and the pessimistic conclusions of Bernstein and Teng (1989) concerning the applicability of decision rules for the number of factors on noncontinuous data, the present study aimed to investigate the performance of parallel analysis in identifying the number of factors on binary data. If parallel analysis performs well for binary data, not only will the results be valuable for researchers applying factor analysis on binary data, but they may also shed light on determining the dimensionality of items constructed within the framework of item response theory.

Method

The present research examined the performance of parallel analysis on unidimensional binary variables. To facilitate comparison with past research, the conditions set out in Parry and McArdle (1991) and Bernstein and Teng (1989) were used as the major references in designing the current study. Single-factor models were simulated with both 8 and 20 indicators. Four independent variables were manipulated, including sample size, factor loading, the proportion of responses on two categories, and the correlation to be analyzed.

Sample sizes were 50, 100, 200, 500, and 1,000. The first three conditions were studied by Parry and McArdle (1991), and the last condition was investigated by Bernstein and Teng (1989). The sample size of 500 was also examined to cover cases that are likely to be encountered in factor analysis (Wang & Weng, 2002). Although the sample size of 50 was included to fully replicate Parry and McArdle's design, factor analysis on such small samples should be discouraged.

The population factor loading was set to be identical for all the variables and included three conditions used by Parry and McArdle (1991): .45, .70, and .90. The chosen factor loadings were close to the values of .50, .71, and .87 used by Bernstein and Teng (1989).

The proportion of responses on two categories had five conditions, as in Parry and McArdle (1991), including 50/50, 60/40, 70/30, 80/20, and 90/10 splits. As Parry and McArdle explained, these values were selected to be representative of real, applied situations. Bernstein and Teng (1989) used 50/50 and 84/16 splits. The first condition was included by Parry and McArdle, and the latter was close to the 80/20 split.

Parallel analysis was conducted on both phi and tetrachoric correlations because all the indicators were dichotomous variables. Tetrachoric correlation was a legitimate method for estimating correlation between the indicators, considering that many dichotomous variables measured in psychological research represented an artificial categorization of underlying continuous latent traits. A cognitive item scored right or wrong reflected to a great extent the result of an examinee's ability compared with a certain threshold on an underlying continuum. Other dichotomous psychological variables measured in clinical or counseling settings might very well share the same property.

There were a total of 150 conditions involved (5 Sample Sizes × 3 Factor Loadings × 5 Splits × 2 Correlations) according to the study design. Bernstein and Teng (1989) reported results on the basis of a single replication. Parry and McArdle (1991) used one replication sample as well, except in certain cases in which 30 replication samples were generated. To evaluate the ratio of correctly identifying the number of factors in all conditions, the present study used 500 replications. SAS (SAS Institute Inc., 1999) was used to generate separate data sets of standard normal variables for each level of sample size. The continuous measures X were created according to the factor analysis model, $X = \text{Loading} \times F + (1 - \text{Loading}^2)^{1/2} \times E$, where F and E are both standard normal variables representing factor scores and error scores, respectively, and the values of loadings varied according to the conditions set forth. The appropriate z values were used as thresholds to transform continuous variables X to dichotomous responses with desired splits as in Bernstein and Teng (1989) and Parry and McArdle (1991).

After the binary data were created, SAS was used to calculate phi correlations and eigenvalues of correlation matrices, and EQS (Bentler, 1995) was used to compute tetrachoric correlations. Parallel analysis was conducted with eigenvalues from simulated sample correlation matrices compared with three criteria. The criteria of mean eigenvalues, 95th-percentile eigenvalues, and 99th-percentile eigenvalues from 1,000 random data correlation matrices were obtained from the SAS/IML program by O'Connor (2000). The number of factors equaled the number of sample eigenvalues greater than the criteria. The ratio of correctly identifying the number of factors as one among the 500 replications was the key element for evaluating the performance of parallel analysis with unidimensional binary data in this study.

Results

Single-Factor Model With Eight Indicators

Because of the extreme distribution of the 90/10 split in the present design, when the sample size was 50, some dichotomous variables turned out to be constant, with all observations falling in the same category, resulting in zero covariance with all other variables. The constant-variable condition happened in 1.8%, 1.8%, and 1.2% of the 500 replications for factor loading of .45, .70, and .90, respectively. In this case, the number of meaningful random variables was reduced to seven. Therefore, the samples were eliminated from analysis and were substituted for by new samples so that each condition would contain a complete set of 500 replications.

Being Gramian is a desired property for any correlation matrix. However, it is well known that matrices of tetrachoric correlation might be non-Gramian (Wothke, 1993). When the sample size was as small as 50, with a factor loading of .90 or with a distribution of 90/10, more than 90% of the matrices using the tetrachoric correlation were non-Gramian. With a sample size of 100 and an extreme distribution of 90/10, the likelihood of obtaining a non-Gramian matrix was still more than 0.80. The likelihood of obtaining a non-Gramian matrix decreased as the sample size increased. None of the matrices was non-Gramian with a sample size of 500 or 1,000.

Effect size η^2 ($SS_{\text{effect}}/SS_{\text{total}}$) was computed to reflect the influences of manipulated variables on the performance of parallel analysis in correctly identifying the number of factors, as presented in Table 1. The variables manipulated had minimal effects on the behaviors of parallel analysis on continuous variables. Only the interaction of factor loading and sample size reached a medium-sized effect (Cohen, 1988). Parallel analysis conducted on eigenvalues of phi correlation matrices was influenced mainly by factor loading and sample size, whereas analysis on eigenvalues of tetrachoric correlation matrices was affected by factor loading and the proportion of responses on two categories. Overall, factor loading had the greatest impact, and the effects of sample size and the distribution of dichotomous variables depended on the type of correlation coefficients analyzed. Sample size affected the performance of parallel analysis on phi correlations, whereas the proportion of responses on two categories affected the behaviors of parallel analysis on tetrachoric correlations.

Tables 2 through 4 present the ratios when parallel analysis correctly identified the number of factors with factor loadings of .90, .70, and .45, respectively. The conditions that reached a 100% correct rate with both phi and tetrachoric correlations are omitted in the tables for simplicity of presentation. Parallel analysis yielded a very accurate decision on the number of factors with the model studied. The larger the sample size, the higher the factor

Table 1
η^2 *of Manipulated Variables on the Correct Decision of the Number of Factors by Parallel Analysis With Eight Indicators*

	Continuous Measures			Phi Correlation			Tetrachoric Correlation		
Criterion	Mean	P95	P99	Mean	P95	P99	Mean	P95	P99
N	.031	.016	.022	.084	.072	.091	.035	.015	.012
L	.031	.010	.012	.130	.074	.094	.200	.090	.070
P				.064	.036	.021	.056	.072	.076
N × L	.063	.031	.043	.059	.076	.125	.034	.018	.016
N × P				.017	.020	.013	.018	.030	.031
L × P				.026	.020	.014	.049	.086	.098
N × L × P				.021	.013	.013	.025	.029	.030
Error	.874	.944	.923	.599	.691	.629	.583	.661	.667

Note. All the effects were statistically significant at $p < .001$. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; N = sample size; L = factor loading; P = proportions of responses on two categories.

Table 2
Correct Rate for Determining the Number of Factors With a Factor Loading of .90 and Eight Indicators

		Phi Correlation			Tetrachoric Correlation		
N	P	Mean	P95	P99	Mean	P95	P99
50	C	1.000	1.000	1.000			
50	0.80	.996	1.000	1.000	.996	1.000	1.000
50	0.90	.830	.904	.928	.984	.996	1.000
100	C	1.000	1.000	1.000			
100	0.90	.966	.988	.990	.994	.998	1.000
200	C	1.000	1.000	1.000			
500	C	1.000	1.000	1.000			
1,000	C	1.000	1.000	1.000			

Note. Conditions with all six entries equal to 1.000 were omitted. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; N = sample size; P = proportions of responses on two categories; C = continuous measures.

loading, and the closer the proportions responding to two categories, the better chance that parallel analysis would result in a correct decision on the number of factors. Only when factor loading was minimum and the sample size was very small did the likelihood of suggesting correct number of factors diminish. Using the 95th- or 99th-percentile eigenvalues as the criteria for comparison yielded a higher correct rate than using the mean eigenvalues. The differences in percentages correctly identifying the number of factors between phi and tetrachoric correlations were minimal. Large differences

Table 3
Correct Rate for Determining the Number of Factors With a Factor Loading of .70 and Eight Indicators

		Phi Correlation			Tetrachoric Correlation		
N	P	Mean	P95	P99	Mean	P95	P99
50	C	1.000	1.000	1.000			
	0.50	.966	.992	.996	.934	.984	.992
	0.60	.952	.994	.998	.920	.974	.992
	0.70	.916	.990	.994	.852	.964	.980
	0.80	.800	.952	.974	.722	.896	.944
	0.90	.390	.702	.786	.828	.954	.976
100	C	1.000	1.000	1.000			
	0.50	.998	1.000	1.000	.986	.998	.998
	0.60	1.000	1.000	1.000	.990	1.000	1.000
	0.70	.998	.998	1.000	.978	.998	.998
	0.80	.940	1.000	1.000	.898	.966	.986
	0.90	.566	.834	.898	.538	.732	.786
200	C	1.000	1.000	1.000			
	0.60	.998	1.000	1.000	.998	.998	.998
	0.80	.998	1.000	1.000	.996	1.000	1.000
	0.90	.872	.970	.984	.778	.876	.914
500	C	1.000	1.000	1.000			
	0.90	.994	.998	1.000	.994	.998	.998
1,000	C	1.000	1.000	1.000			

Note. Conditions with all six entries equal to 1.000 were omitted. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; N = sample size; P = proportions of responses on two categories; C = continuous measures.

usually occurred when the two response categories were of an extreme split, such as 90/10.

Continuous measures. As shown in Tables 2 through 4, parallel analysis with continuous measures performed very well regardless of the criterion used as the baseline for comparison. The lowest correct rate was 0.830, when the sample size was 50 and factor loading was .45, with mean eigenvalues as the criterion.

Factor loading of .90. Table 2 shows the percentage of replications for which parallel analysis correctly identified the number of factors at a factor loading of .90. With factor loading as high as .90, parallel analysis using both phi and tetrachoric correlations yielded an almost perfect correct rate of determining the number of factors, even when the sample size was small and the split of responses on two categories was extremely uneven. With a sample size of 50 and a 90/10 split, phi correlations gave a slightly lower correct rate

Table 4
Correct Rate for Determining the Number of Factors With a Factor Loading of .45 and Eight Indicators

		Phi Correlation			Tetrachoric Correlation		
N	P	Mean	P95	P99	Mean	P95	P99
50	C	.830	.934	.914			
	0.50	.568	.704	.570	.702	.958	.992
	0.60	.542	.672	.548	.702	.942	.980
	0.70	.512	.634	.556	.628	.926	.966
	0.80	.376	.506	.430	.336	.686	.808
	0.90	.238	.378	.356	.638	.872	.906
100	C	.942	.994	.996			
	0.50	.734	.922	.906	.660	.904	.954
	0.60	.700	.900	.894	.640	.904	.948
	0.70	.624	.878	.840	.548	.846	.918
	0.80	.512	.772	.722	.466	.800	.874
	0.90	.272	.514	.486	.062	.226	.298
200	C	.996	1.000	1.000			
	0.50	.906	.992	.996	.720	.924	.966
	0.60	.882	.986	.992	.708	.918	.944
	0.70	.836	.984	.996	.624	.870	.918
	0.80	.704	.928	.952	.444	.762	.820
	0.90	.312	.676	.726	.146	.322	.394
500	C	1.000	1.000	1.000			
	0.50	.996	1.000	1.000	.950	.992	.996
	0.60	.992	1.000	1.000	.936	.984	.990
	0.70	.984	.994	.998	.868	.958	.974
	0.80	.908	.988	.996	.638	.850	.882
	0.90	.612	.888	.958	.234	.460	.538
1,000	C	1.000	1.000	1.000			
	0.50	1.000	1.000	1.000	.996	1.000	1.000
	0.60	1.000	1.000	1.000	.994	1.000	1.000
	0.70	1.000	1.000	1.000	.988	1.000	1.000
	0.80	.998	1.000	1.000	.940	.982	.990
	0.90	.856	.982	.992	.486	.678	.734

Note. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; N = sample size; P = proportions of responses on two categories; C = continuous measures.

than tetrachoric correlations, and the performance of mean eigenvalues was worse than the 95th- or 99th-percentile eigenvalues.

Factor loading of .70. Table 3 presents the ratio for which parallel analysis correctly identified the number of factors with a factor loading of .70. With a factor loading of .70, parallel analysis that used the 95th-percentile eigenvalues and the 99th-percentile eigenvalues of random data correlation matrices as the criteria correctly determined the number of factors, except for the case of a 90/10 split. Phi correlations and tetrachoric correlations gave similar results. However, phi correlations performed better than tetrachoric correlations in certain cases with a sample size less than 500. When the sample size was 50, the mean eigenvalues gave a lower correct rate than the 95th- and 99th-percentile eigenvalues, and tetrachoric correlations performed worse than phi correlations, except for the case of the 90/10 split. With a 90/10 split, the correct rate of phi correlations was either higher than or equal to that of tetrachoric correlations, except for the case of a very small sample size of 50.

Factor loading of .45. Table 4 reports the percentage of replications that correctly identified the number of factors for a factor loading of .45. When the factor loading was .45, the performance of mean eigenvalues was again worse than the 95th- and 99th-percentile eigenvalues in identifying the correct number of factors. Phi correlations outperformed tetrachoric correlations, except with very small samples. With an extreme split of 90/10, parallel analysis performed poorly. As observed from Table 4, with low factor loadings, parallel analysis was likely to suggest an incorrect number of factors when sample size was small and percentages responding to two categories varied widely. For instance, if a correct rate of .90 was aimed for with phi correlations subject to parallel analysis, a sample size of at least 200 was required, and the split of responses on two categories could not be as extreme as 90/10.

Single-Factor Model With 20 Indicators

Constant variables again occurred with the extreme distribution of a 90/10 split and a small sample size of 50. Such cases were found in 4.2%, 3.8%, and 2.8% of the 500 replications with factor loading of .45, .70, and .90, respectively, and were replaced with new samples. The likelihood of obtaining a non-Gramian matrix using a tetrachoric correlation was more than 0.95 when the sample size was as small as 50. The likelihood approached 1.00 for a sample size of 100, except when a factor loading of .45 was combined with more symmetric distributions. When the sample size was 200, the likelihood of obtaining a non-Gramian matrix with a tetrachoric correlation was greater than 0.98 if the factor loading was .90 or if the distribution was a 90/10 split. Although the probability of a non-Gramian matrix declined as the sample size reached 500, more than 98% of the matrices using the tetrachoric correlation under the combination of a high factor loading (.90) and extreme distribution (a 90/10 split) were non-Gramian. Overall, when a tetrachoric correlation was calculated, the 20-indicator model was more likely to yield non-Gramian correlation matrices than the 8-indicator model.

Table 5
η^2 *of Manipulated Variables on the Correct Decision of the Number of Factors by Parallel Analysis With 20 Indicators*

	Continuous Measures			Phi Correlation			Tetrachoric Correlation		
Criterion	Mean	P95	P99	Mean	P95	P99	Mean	P95	P99
N	.018	.002	.000	.093	.053	.044	.043	.023	.022
L	.014	.001	.000	.116	.035	.021	.392	.241	.191
P				.184	.176	.148	.082	.140	.158
N × L	.037	.004	.001	.033	.012	.009	.048	.035	.036
N × P				.048	.096	.103	.008	.016	.022
L × P				.051	.051	.040	.040	.100	.122
N × L × P				.045	.026	.022	.051	.046	.052
Error	.931	.992	.999	.430	.550	.613	.337	.399	.398

Note. All the effects were statistically significant at $p < .001$. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; N = sample size; L = factor loading; P = proportions of responses on two categories.

Table 5 presents the effect sizes of manipulated variables on the performance of parallel analysis on single-factor model with 20 indicators in correctly determining the number of factors. Again, the manipulated variables had a very limited impact with continuous measures. When binary measures were analyzed, the performance of phi correlations was mainly influenced by the proportion of responses on two categories, followed by factor loading, whereas the behaviors of tetrachoric correlations were largely determined by factor loading, followed by response proportions on two categories. Compared with the model of eight indicators, the influence of the proportion of responses on two categories increased when phi correlations among 20 indicators were analyzed, and the effect size of factor loading doubled when tetrachoric correlations were used.

Tables 6 through 8 present the percentage of analyses in which the number of factors was correctly identified by parallel analysis over the 500 replications at each level of factor loading. The conditions that reached a 100% correct rate with both phi and tetrachoric correlations are again omitted in the tables for simplicity of presentation. The first row at each level of sample size reports the performance of parallel analysis against continuous measures. Parallel analysis again performed very well on continuous measures, with the lowest correct rate of 0.913 for a sample size of 50 and a factor loading of .45 using mean eigenvalues as the basis for comparison. With binary data, the 95th- and 99th-percentile eigenvalues again outperformed mean eigenvalues, and phi correlations demonstrated a better chance of correctly identifying the number of factors than tetrachoric correlations. In general, closer proportions of responses on two categories, higher factor loading, and

Table 6
Correct Rate for Determining the Number of Factors With a Factor Loading of .90 and 20 Indicators

		Phi Correlation			Tetrachoric Correlation		
N	P	Mean	P95	P99	Mean	P95	P99
50	C	1.000	1.000	1.000			
50	0.80	.983	.998	1.000	.980	.992	.997
50	0.90	.492	.652	.710	.833	.915	.940
100	C	1.000	1.000	1.000			
100	0.90	.855	.927	.955	.943	.980	.985
200	C	1.000	1.000	1.000			
200	0.90	.985	.992	.993	.993	.997	.998
500	C	1.000	1.000	1.000			
1,000	C	1.000	1.000	1.000			

Note. Conditions with all six entries equal to 1.000 were omitted. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; N = sample size; P = proportions of responses on two categories; C = continuous measures.

larger sample size would improve the performance of parallel analysis on a single-factor model with 20 binary variables. Exceptions were to be presented in details. However, compared with the model with 8 indicators, the performance of parallel analysis against binary data of extreme distributions became much worse when the number of indicators was increased to 20.

Factor loading of .90. Parallel analysis with 20 highly saturated variables performed well, except for the combination of a very small sample size and an extreme distribution. The correct rate dropped to 0.492 for the case of 50 observations and a 90/10 split when mean eigenvalues from phi correlations were used as the criterion for comparison.

Factor loading of .70. The performance of the 95th- and 99th-percentile eigenvalues was similar and yielded a better correct rate than mean eigenvalues, as in the model with 8 indicators. Phi correlations performed better than tetrachoric correlations, except when the sample size was 50 and the distribution was a 90/10 split. In general, for single-factor models with a .70 factor loading, parallel analysis with 20 indicators performed worse than with only 8 indicators.

Factor loading of .45. The performance of mean eigenvalues was again worse than the 95th- and 99th-percentile eigenvalues in correctly deciding the number of factors. Phi correlations performed better than tetrachoric correlations in most cases. When tetrachoric correlations were analyzed, the sample size and the correct rate displayed a U-shaped relation, with a sample

Table 7
Correct Rate for Determining the Number of Factors With a Factor Loading of .70 and 20 Indicators

		Phi Correlation			Tetrachoric Correlation		
N	P	Mean	P95	P99	Mean	P95	P99
50	C	1.000	1.000	1.000			
	0.50	.988	1.000	1.000	.803	.968	.982
	0.60	.983	.998	1.000	.792	.955	.978
	0.70	.943	.995	.998	.673	.888	.937
	0.80	.673	.907	.957	.325	.603	.737
	0.90	.067	.242	.353	.387	.653	.747
100	C	1.000	1.000	1.000			
	0.50	.998	1.000	1.000	.973	.993	.997
	0.60	1.000	1.000	1.000	.967	.998	1.000
	0.70	.995	1.000	1.000	.858	.957	.980
	0.80	.940	.992	.997	.562	.770	.850
	0.90	.188	.468	.595	.052	.135	.195
200	C	1.000	1.000	1.000			
	0.60	1.000	1.000	1.000	.998	.998	.998
	0.70	1.000	1.000	1.000	.997	1.000	1.000
	0.80	.995	.995	.998	.920	.965	.982
	0.90	.558	.820	.857	.227	.388	.417
500	C	1.000	1.000	1.000			
	0.90	.980	.998	1.000	.878	.940	.960
1,000	C	1.000	1.000	1.000			
	0.90	1.000	1.000	1.000	.998	1.000	1.000

Note. Conditions with all six entries equal to 1.000 were omitted. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; N = sample size; P = proportions of responses on two categories; C = continuous measures.

size of 200 behaving the worst. Another result worth noting was that when the distributions of binary measures were extreme, the correct rate was very low. In this case, increasing the sample size helped bring up the chance of correctly identifying the number of factors for phi correlations, but increasing the sample size did not necessarily raise the correct rate if tetrachoric correlations were subject to analysis.

Discussion

The applicability of parallel analysis against unidimensional binary data was investigated in the present study. The results of the simulation study suggested that parallel analysis performed well in identifying the correct number of factors under most conditions. Bernstein and Teng (1989) suggested that the criteria for identifying number of factors applicable to continuous data were inappropriate for discrete data. However, on the basis of the findings of

Table 8
Correct Rate for Determining the Number of Factors With a Factor Loading of .45 and 20 Indicators

		Phi Correlation			Tetrachoric Correlation		
N	P	Mean	P95	P99	Mean	P95	P99
50	C	.913	.992	.998			
	0.50	.687	.945	.960	.613	.930	.977
	0.60	.605	.933	.965	.630	.935	.975
	0.70	.510	.882	.927	.467	.863	.955
	0.80	.248	.680	.815	.038	.253	.455
	0.90	.042	.242	.350	.245	.513	.663
100	C	.983	1.000	1.000			
	0.50	.822	.962	.992	.308	.747	.870
	0.60	.797	.970	.995	.310	.745	.870
	0.70	.622	.935	.982	.175	.572	.745
	0.80	.367	.825	.922	.033	.265	.442
	0.90	.048	.300	.482	.000	.000	.000
200	C	1.000	1.000	1.000			
	0.50	.937	.995	.998	.263	.675	.767
	0.60	.920	.990	.995	.203	.583	.703
	0.70	.875	.985	.993	.087	.397	.505
	0.80	.628	.923	.955	.008	.098	.183
	0.90	.110	.470	.608	.000	.000	.000
500	C	1.000	1.000	1.000			
	0.50	.998	1.000	1.000	.645	.902	.937
	0.60	1.000	1.000	1.000	.590	.842	.927
	0.70	.993	1.000	1.000	.293	.628	.763
	0.80	.913	.990	.998	.032	.147	.248
	0.90	.325	.697	.825	.000	.000	.000
1,000	C	1.000	1.000	1.000			
	0.50	1.000	1.000	1.000	.963	.992	.998
	0.60	1.000	1.000	1.000	.933	.987	.993
	0.70	1.000	1.000	1.000	.820	.938	.973
	0.80	.993	1.000	1.000	.315	.567	.688
	0.90	.645	.892	.953	.000	.002	.007

Note. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; N = sample size; P = proportions of responses on two categories; C = continuous measures.

the current investigation, their conclusion may need some qualification. The present study demonstrated that some criteria designed for continuous variables are appropriate for binary data. Parallel analysis correctly determined the number of factors on single-factor models with 8 and 20 binary indicators in the majority of conditions studied. Poor performance occurred under the combination of a small factor loading and an extreme distribution of responses on two categories. A careful examination of the scale properties of the measures investigated and the adoption of appropriate methods for analysis are indispensable research practices. A researcher who fails to use parallel analysis for binary data may falsely conclude unidimensional data to be multidimensional.

Parallel analysis using the 95th- and 99th-percentile eigenvalues as the basis for determining the number of factors outperformed that using the criterion of mean eigenvalues. The results coincided with those obtained in previous research (e.g., Glorfeld, 1995; Turner, 1998). We therefore suggest that either the 95th- or the 99th-percentile eigenvalues be used for future applications of parallel analysis in determining the number of factors.

However, could it be possible that the better performance of the 95th- and 99th-percentile eigenvalues or the general excellent performance of parallel analysis comes at the expense of underextracting the number of factors such that multidimensional data might be incorrectly identified as unidimensional? Parallel analysis under the present design of a unidimensional model is not likely to extract too few factors and would mainly err on the side of extracting too many factors. Multidimensional data are needed to assess whether parallel analysis or the strategy of using the 95th- and 99th-percentile eigenvalues is so stringent that it may tend to identify a smaller number of factors in multidimensional data. Additional simulations with multidimensional data were therefore conducted to assess whether parallel analysis underextracts the number of factors.

An orthogonal two-factor model with eight indicators on each factor was used with the same four independent variables as in previous simulations manipulated. The factor loading, proportion of responses on two categories, and correlation to be analyzed included the same conditions as in the main study. Without a loss of generality, the sample size was evaluated at 200 and 500 for the sake of simplicity. Each condition included 500 replication samples. The results of the two-factor model followed the general conclusions from the one-factor model that parallel analysis correctly determined the number of factors and analysis using the 95th- and the 99th-percentile eigenvalues of the criteria that outperformed using mean eigenvalues as the basis for comparison. Moreover, when parallel analysis gave an incorrect number of factors, it was more likely to extract too many factors than to extract too few. The likelihood that the excellent performance of parallel analysis or the 95th- and 99th-percentile eigenvalues comes at the expense of underextraction is minimal. With a factor loading of .70 or .90, parallel analysis yielded an almost 100% correct rate for determining the number of factors. Exceptions usually occurred for variables with the 90/10 split. Table 9 therefore presents only the percentages of underextraction and overextraction of the two-factor model with a factor loading of .45 together with a 90/10 split for illustrative purpose. The results indicated that parallel analysis tended to extract too many factors, especially when tetrachoric cor-

Table 9
Percentages Underextraction (U) and Overextraction (O) for the Two-Factor Model With a Factor Loading of .45

	Phi Correlation						Tetrachoric Correlation					
	Mean		P95		P99		Mean		P95		P99	
P	U	O	U	O	U	O	U	O	U	O	U	O
n = 200												
0.50	0	24.4	.2	3.6	.4	1.0	0	32.0	.6	9.2	1.0	5.0
0.60	0	22.2	.4	3.6	.8	.4	.4	33.8	1.2	8.4	2.2	3.2
0.70	0	28.4	.4	5.4	1.0	2.2	.2	52.6	1.8	17.8	3.8	9.8
0.80	.2	47.4	2.6	11.4	6.4	3.8	0	85.8	4.2	50.4	7.0	35.0
0.90	1.8	77.2	14.2	29.4	30.2	14.8	0	100.0	0	100.0	0	99.6
0.90[a]	0	12.6	0	3.0	0	2.0	0	92.0	0	87.4	0	84.2
0.90[b]	0	0	0	0	0	0	0	11.6	0	8.4	0	6.8
n = 500												
0.50	0	1.6	0	.2	0	.2	0	6.4	0	1.2	0	.8
0.60	0	3.6	0	.8	0	.6	0	7.6	0	1.0	0	1.0
0.70	0	6.2	0	.2	0	.2	0	19.0	0	4.6	0	2.6
0.80	0	13.0	0	2.4	0	0	0	60.8	0	29.6	0	19.8
0.90	0	57.2	.2	23.4	.8	13.4	0	99.8	0	97.2	0	95.2
0.90[a]	0	.4	0	0	0	0	0	16.0	0	10.6	0	7.4
0.90[b]	0	0	0	0	0	0	0	0	0	0	0	0

Note. Mean = mean eigenvalues of 1,000 random correlation matrices; P95 = 95th-percentile eigenvalues of 1,000 random correlation matrices; P99 = 99th-percentile eigenvalues of 1,000 random correlation matrices; P = proportions of responses on two categories.
a. Factor loading = .70.
b. Factor loading = .90.

relation was subject to analysis and as the proportions of responses on two categories became more uneven.

Phi correlations yielded better performance than tetrachoric correlations in many conditions. A closer examination of the estimation of population correlations by phi and tetrachoric correlations indicated that the sampling error of tetrachoric correlations could be relatively large with small samples, especially when the proportions of responses on two categories varied widely. Consider the eight-indicator model with a factor loading of .45 as an example. In this case, the population correlation was 0.2025. Consider the situation in which the sample size was 50 and the responses on the dichotomy were of a 90/10 split. The mean tetrachoric correlation over the 500 replications was .537, with a standard error of .216. The behaviors of tetrachoric correlations were rather unstable with a sample size less than 200. Therefore, the tetrachoric correlation that may be theoretically more legitimate in dealing with artificially dichotomized measures than phi correlation does not necessarily lead to satisfactory outcomes. Parry and McArdle (1991) demonstrated the similar phenomena in estimating factor loadings of binary items by phi and tetrachoric correlations. Tetrachoric correlation was not found to be superior to phi correlation. Future applications of parallel analysis on binary data could be carried out with phi correlations instead of tetrachoric correlations in view of the unstable behavior of tetrachoric correlation.

The size of factor loading played a critical role in the performance of parallel analysis, as did the proportions of responses on two categories. The higher the factor loading or the closer the proportion of responses on each category, the better chance that parallel analysis would indicate the correct number of factors. Between these two essential elements for the performance of parallel analysis, researchers may have more control over factor loadings than participant responses on scales. Some psychometric measures would by nature show highly disproportional responses on the absence or presence of symptoms. Researchers should invest their endeavor in writing quality items that are representative of and highly related to the construct being evaluated. Moreover, items of high loadings not only avoid errors in estimating number of factors but also improve the quality of factor loading estimation (MacCallum, Widaman, Zhang, & Hong, 1999; Velicer & Fava, 1998).

Parallel analysis with larger samples tended to yield a better decision on the number of factors. However, when a small factor loading was combined with an uneven response pattern on dichotomous measures, a very large sample was required for phi correlation to reach satisfactory results, and yet increasing the sample size did not necessarily help correctly identify the number of factors with tetrachoric correlations. A large sample size was unable to compensate for the poor performance of parallel analysis because of small loadings and extreme distributions.

The present simulations examined the performance of parallel analysis on single-factor models with identically distributed binary measures. In future investigations, comprehensive multiple-factor models or variables of mixed types of distributions as well as responses of Likert-type ratings can be considered to fully understand the applicability of parallel analysis on various conditions encountered in educational and psychological research in the real world. The procedure of parallel analysis has been modified to assess the dimensionality of items constructed by unidimensional item response theories (Budescu, Cohen, & Ben-Simon, 1997; Drasgow & Lissak, 1983). The modified procedures require the generation of random data on the basis of estimated item parameters and are more complicated than the original method of parallel analysis. Compared with the complex procedures involved in modified parallel analysis, the possibility of parallel analysis on examining the dimensionality of item response theory items also seems worth investigating.

References

Allen, S. J., & Hubbard, R. (1986). Regression equations for the latent roots of random data correlation matrices with unities on the diagonal. *Multivariate Behavioral Research*, *21*, 393-398.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, *105*, 467-477.

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.

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

Collins, L. M., Cliff, N., McCormick, D. J., & Zatkin, J. L. (1986). Factor recovery in binary data sets: A simulation. *Multivariate Behavioral Research*, *21*, 377-391.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

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.

Drasgow, 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.

Enzmann, D. (1997). RanEigen: A program to determine the parallel analysis criterion for the number of principal components. *Applied Psychological Measurement*, *21*, 232.

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 application 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.

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.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 1179-1185.

Humphreys, L. G., & Ilgen, D. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement*, *29*, 571-578.

Humphreys, L. G., & Montanelli, R. G., Jr. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, *10*, 193-205.

Kaufman, J. D., & Dunlap, W. P. (2000). Determining the number of factors to retain: A Windows-based FORTRAN-IMSL program for parallel analysis. *Behavior Research Methods, Instruments, and Computers*, *32*, 389-395.

Keeling, K. B. (2000). A regression equation for determining the dimensionality of data. *Multivariate Behavioral Research*, *35*, 457-468.

Lautenschlager, G. J. (1989a). A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research*, *24*, 365-395

Lautenschlager, G. J. (1989b). PARANAL.TOK: A program for developing parallel analysis criteria. *Applied Psychological Measurement*, *13*, 176.

Lautenschlager, G. J., Lance, C. E., & Flaherty, V. L. (1989). Parallel analysis criteria: Revised equations for estimating the latent roots of random data correlation matrices. *Educational and Psychological Measurement*, *49*, 339-345.

Longman, R. S., Cota, A. A., Holden, R. R., & Fekken, G. C. (1989a). PAM: A double-precision FORTRAN routine for the parallel analysis method in principal components analysis. *Behavior Research Methods, Instruments, and Computers*, *21*, 477-480.

Longman, R. S., Cota, A. A., Holden, R. R., & Fekken, G. C. (1989b). A regression equation for the parallel analysis criterion in principal components analysis: Mean and 95th percentile eigenvalues. *Multivariate Behavioral Research*, *24*, 59-69.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99.

Montanelli, R. G., Jr., & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika*, *41*, 341-348.

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, and Computers*, *32*, 396-402.

Parry, C. D. H., & McArdle, J. J. (1991). An applied comparison of methods for least-squares factor analysis of dichotomous variables. *Applied Psychological Measurement*, *15*, 35-46.

Reilly, A., & Eaves, R. C. (2000). Factor analysis of the Minnesota Infant Development Inventory based on a Hispanic migrant population. *Educational and Psychological Measurement*, *60*, 271-285.

Roznowski, M., Tucker, L. R., & Humphreys, L. G. (1991). Three approaches to determining the dimensionality of binary items. *Applied Psychological Measurement*, *15*, 109-127.

Sarff, P. L. (1997). Using factor replicability to evaluate number of factors rules. *Dissertation Abstracts International, Section B: The Sciences & Engineering*, *58(4-B)*, 2175.

SAS Institute Inc. (1999). *SAS/IML user's guide, Version 8*. Cary, NC: Author.

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

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 N. Jackson at seventy* (pp. 41-71). Norwell, MA: Kluwer Academic.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*, 231-251.

Wang, C.-N. (2001). *Effect of number of response categories and scale distribution on factor analysis*. Unpublished master's thesis, National Taiwan University, Taipei, Taiwan.

Wang, C.-N., & Weng, L.-J. (2002). Evaluating the use of exploratory factor analysis in Taiwan: 1993-1999. *Chinese Journal of Psychology*, *44*, 239-251.

Weng, L.-J. (1995). An overview of applications of factor analysis. In Y.-H. Chang, Y.-C. Fu, & H.-Y. Chiu (Eds.), *Social survey and analysis: The first interdisciplinary symposium on research methods in the social sciences* (pp. 245-265). Taipei, Taiwan: Institute of Ethnology, Academia Sinica.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256-293). Newbury Park, CA: Sage.

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