

A Comparison of Methods for Dimensionality Assessment of Categorical Item Responses

Chen-Wen Liu and Wen-Chung Wang

Exploratory factor analysis (EFA) and principal component analysis (PCA) are two common methods for dimensionality assessment. Both methods assume that the variables are continuous. In EFA, a few common factors are extracted from the variables; whereas in PCA, a few components are created to account for the variations among variables. Linear EFA or PCA was enhanced by means of parallel analysis (PA). In the PA method, a large number (e.g., 100) of datasets that have the same size of response matrix of the real dataset are randomly simulated and then analyzed with linear EFA or PCA. These eigenvalues obtained from random datasets are compared with those of the real dataset. The number of factors is the number of times when the eigenvalues derived from the real dataset are larger than the 95th percentile (or mean) of the eigenvalue distribution of the simulated datasets. The PA method appears promising in dimensionality assessment for continuous, dichotomous, and polytomous variables, and it can accommodate the Pearson correlation and the polychoric correlation (Cho et al. 2009; Tran and Formann 2009; Weng and Cheng 2005).

Weng and Cheng (2005) concluded that the PA method performs well for unidimensional dichotomous variables when the 95th or 99th percentile of the random data eigenvalues criteria is used. Cho et al. (2009) observed that the PA method based on the Pearson correlation performs at least as well as that based on the polychoric correlation in most conditions. However, Tran and Formann (2009) showed that the PA method does not perform well in assessing the dimensionality of dichotomous variables based on the Pearson correlation or the tetrachoric correlation. Finch and Monahan (2008) implemented a modified PA method based on nonlinear factor analysis with the TESTFACT program (Wilson et al. 1991) and

C.-W. Liu (✉) · W.-C. Wang
Education University of Hong Kong, Tai Po, Hong Kong
e-mail: cwliu@eduhk.hk

W.-C. Wang
e-mail: wcwang@eduhk.hk

parametric bootstrap sampling, and they concluded that the new method outperforms the DIMTEST method in identifying the unidimensional structure. Timmerman and Lorenzo-Seva (2011) implemented the PA method with a minimum rank factor analysis to assess the dimensionality of polytomous variables, and they observed that this method outperforms traditional PA methods in most conditions. They thus recommended the use of the polychoric correlation for polytomous variables if there is no convergence problem. In this study, the PA method implemented in the computer program FACTOR (Lorenzo-Seva and Ferrando 2006) was used to serve as a baseline, with which other methods were compared.

The Hull method can be regarded as a generalization of the scree test (Lorenzo-Seva et al. 2011). A series of EFAs are conducted, each with a different number of factors, starting from 1 to an upper bound. The upper bound can be determined by the indicated number of factors via the PA method plus one for the search range of the elbow on the scree plot. Each run has several goodness-of-fit indices (e.g., the comparative fit index, root mean square error of approximation, standardized root mean square residual, and common part accounted for) which are plotted against the degrees of freedom in a two-dimensional scree-like plot. The Hull method seeks an optimal balance between the fit and the degrees of freedom iteratively. The number of dimensions is determined, in which the elbow is located heuristically on a break or discontinuity in the convex hull. Lorenzo-Seva et al. (2011) conducted simulation studies on continuous variables to compare the Hull method and the PA method using the Bayesian information criterion, and observed that the Hull method outperforms the PA method in recovering the correct number of dimensions.

The DETECT method is a statistical method for dimensionality assessment. It aims to identify not only the number of dimensions, but also by which items a dimension is predominantly measured. The DETECT method is based on covariance theory that items measuring the same dimension would have positive covariances, whereas items measuring different dimensions would have negative covariances, given the latent traits or total scores have been taken into account (Stout et al. 1996; Zhang 2007; Zhang and Stout 1999). The DETECT index is expected to be zero if data are truly unidimensional. If data are multidimensional, the expected pairwise conditional covariance will be positive when both items measure the same dimension, and negative when they measure different dimensions. At the outset, a genetic algorithm, together with a hierarchical cluster analysis, can be used to search an optimal subspace. The optimal dimensionality partition is obtained by searching all over possible dimensions to maximize the DETECT index. The approximate simple structure index and the ratio index can help determine whether a dataset displays an approximately simple structure (in a simple structure, each item measures a single dimension, but different items may measure different dimensions).

The spectral clustering (SC) method aims to classify data structure into clusters on a manifold space (Luxburg 2007). It is widely used to cluster image data or pattern representation in computer science, statistics, etc. Moreover, it is easy to implement via standard linear algebra methods and usually outperforms traditional approaches such as the k -means clustering algorithm (Luxburg 2007). Once the number of clusters (dimensions) is determined, the SC method continues to classify

variables (items) to different clusters. Analogous to the DETECT method, the SC method identifies not only the number of clusters, but also the cluster to which an item belongs. Although the SC method is promising, to the authors' knowledge, it has never been applied to dimensionality assessment within the IRT framework. In this study, we thus evaluated the performance of an extended version of the SC method within the IRT framework. The details of the SC method and its extended version are given in the following section.

To sum up, this study aimed to compare the PA, DETECT, Hull, and SC methods in dimensionality assessment of IRT data through simulations. Most IRT models, including the one-, two-, and three-parameter logistic models for dichotomous items (Birnbaum 1968; Rasch 1960), and the partial credit model (Masters 1982), generalized partial credit model (Muraki 1992), and graded response model for polytomous items were examined. The rest of this article is organized as follows. First, we introduce the background of the SC method and its variant of spectral multi-manifold clustering (SMMC), which is rarely (or never) used in the IRT field. Second, we describe the IRT models that were used for data generation. Third, we summarize the results of the simulations that were conducted to evaluate the performance of the PA, DETECT, Hull, and SC methods in dimensionality assessment. Finally, we draw some conclusions and make suggestions for future studies.

Spectral Multi-manifold Clustering (SMMC)

In machine-learning terminology, a low-dimensional manifold is a topological space hidden in a high-dimensional data space. Conceptually, each item is located in an N -dimensional space, and items measuring the same construct should form a one-dimensional manifold. Trying to discover the dimensionality of item response data is intuitively analogous to discovering their latent manifolds. The characteristics of the *similar* items are learned by a machine-learning algorithm based on the responses. The properties of a high-dimensional data space are preserved in a low-dimensional embedding of the data, a process known as *dimensionality reduction*. After feature extraction, the next step is to classify items into appropriate clusters in the low-dimensional data space. The two-step processes are carried out consecutively in the SC method.

In essence, the SC method starts with generating an undirected affinity matrix $\mathbf{G}(V, E)$, which is derived from pairwise similarities (e.g., Euclidean distance). The term "undirected" means an equal weight between two data points, regardless of their direction. V and E stand for vertices (i.e., data points) and edge (i.e., the weight distance between vertices), respectively. The affinity matrix can be constructed by the ε -neighborhood graph, the k -nearest neighbor graphs, or the fully connected graph. The choice of construction depends on the problems to be solved. First, one can select the k -nearest neighbor graphs where a data point connects to only k -nearest data points near itself, finally resulting in a sparse, undirected, and

weighted adjacency matrix \mathbf{W} (Luxburg 2007). Second, the normalized SC algorithm can be used (Shi and Malik 2000), which meets the objectives of clustering (i.e., minimize the between-cluster similarities and maximize the within-cluster similarities), consistence (i.e., the clustering results converge to a true partition when data increase), and simple computation (Luxburg 2007). Therefore, the data points can be mapped into a low-dimensional space, where the characteristics of sets of items become more evident than in the original space (i.e., similar items are close to one another than to other items). Finally, the number of dimensions is determined by counting the first k smallest eigenvalues equal to zero, and then the traditional k -means algorithm can be used to cluster an eigenvector matrix derived from a generalized eigenproblem. Items within the same cluster are deemed as measuring the same dimension.

The normalized SC algorithm is outlined as follows. First, a similarity graph is constructed using the k -nearest neighbor method (or the other methods outlined above). Second, an unnormalized graph Laplacian matrix is computed, which is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where \mathbf{D} is a degree matrix, defined as the diagonal matrix with the degrees d_1, \dots, d_l on the diagonal. The d_i is defined as. Third, a generalized eigenproblem $\mathbf{L}u = \lambda\mathbf{D}u$ is solved to derive the first k generalized eigenvector u_1, \dots, u_k corresponding to the k smallest eigenvalues, then a matrix $\mathbf{U} \in \mathbb{R}^{l \times k}$ is obtained, where k is the number of intrinsic dimension. The element of \mathbf{U} is $y_i \in \mathbb{R}^k$ for $i = 1, \dots, l$. Fourth, the items are grouped by the k -means algorithm into clusters.

In a multidimensional space, we conjecture that items in the same group (i.e., measuring the same dimension) may intersect with, or be nearly close to, other groups of items, especially when these dimensions are highly correlated. However, the normalized SC algorithm merely includes a local similarity among points in the neighborhood. It may not be informative to accurately cluster items. Thus, the structural similarity (i.e., similar local tangent space) is another source of information derived from the data. Wang et al. (2011) proposed the SMMC algorithm to consider the information of local similarity and structural similarity to compositely create an affinity matrix \mathbf{W} . The main idea is that if two items are close to each other and have similar tangent spaces, they may come from the same manifold (dimension). Then, local tangent space i at an item i can be approximated mixtures of probabilistic principal component analyzers (Tipping and Bishop 1999). The tangent space of pairwise points can be defined as, where h is the number of dimensions of the manifolds; ω is a weighting parameter; and form the principal angles between two tangent spaces and. The is defined as where, $l = 2, \dots, h$.

The local similarity s_{ij} is defined as

$$s_{ij} = \begin{cases} 1 & \text{if } x_i \in \text{Knn}(x_j) \text{ or } x_j \in \text{Knn}(x_i) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $\text{Knn}(x)$ denotes k -nearest neighbors of x . Finally, the local similarity function and structural similarity function are integrated to generate the affinity value:

$$w_{ij} = s_{ij}t_{ij} = \begin{cases} \left[\sum_{l=1}^h \cos(\phi_l) \right]^\omega & \text{if } x_i \in Knn(x_j) \text{ or } x_j \in Knn(x_i) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Each item response vector is transformed into a standardized score to remove the impact of different variances of response vector in calculating the Euclidean distance between items.

Overall, the SMMC method uses only the distance information of local similarity function and structural similarity of the items, and does not assume the underlying item response function. More item categories could reduce the measurement error in calculating the Euclidean distances between items (i.e., the ordinal scale could be approximated as an interval scale as the number of categories increases), which could increase the accuracy. Sample size could also be an important factor and were investigated in the simulation studies.

Multidimensional IRT Models

The multidimensional generalized partial credit model (MGPCM) (Yao and Schwarz 2006) was used to generate item responses. It is commonly used for achievement tests in the psychometric literature. The item response function of the MGPCM is defined as

$$\Pr(Z_{ni} = z | \theta_n, \xi_i) = \frac{\exp[z(\alpha_i^T \theta_n - \delta_i) - \sum_{k=0}^z \tau_{ik}]}{\sum_{m=0}^{C-1} \{ \exp[m(\alpha_i^T \theta_n - \delta_i) - \sum_{k=0}^m \tau_{ik}] \}}, \quad (3)$$

where $Z_i \in [0, 1, \dots, C - 1]$ is the observable response variable; z is the observed response; C is the number of categories; θ_n is the latent trait vector representing $D \times 1$ dimensions for person, n ; ξ_i is item i 's parameter vector containing a $D \times 1$ slope parameter vector of α_i , an intercept parameter of δ_i ; and threshold parameters of $\tau_{i0}, \dots, \tau_{iC-1}$. τ_{i0} is set at 0 by convention. The MGPCM is compensatory because the latent traits θ are connected by a linear combination of α_i , when an item measures multiple latent traits. Dimensionality is the minimum number of latent traits that is adequate to explain the underlying examinees' performance (assuming local independence and monotonicity) and groups of items being sensitive to differences along the dimensions (Reckase 2009; Svetina and Levy 2014).

To be in line with most literature on dimensionality assessment, we focus on simple structures of latent traits in this study. That is, each item measures a single dimension and different items may measure different latent traits. Simple structures are also referred to as between-item multidimensionality, in contract to within-item multidimensionality where an item may measure multiple latent traits simultaneously (Adams et al. 1997).

Simulations

The SMMC, PA, and Hull methods have been implemented and are available at http://lamda.nju.edu.cn/code_SMMC.ashx (Wang et al. 2011) and <http://psico.fcep.urv.es/utilitats/factor/>, respectively, where the last two methods have been implemented in a stand-alone program called FACTOR 8.1 (Lorenzo-Seva and Ferrando 2006). The DOS version of the poly-DETECT program is generously provided by the author (Zhang 2007).

In this study, the independent variables included: (a) number of respondents ($N = 250; 1000; 2000$), (b) number of response categories ($C = 2, 4, 6$), (c) number of dimensions ($D = 1, 2, 3$), (d) correlation among dimensions ($r = 0, 0.4, 0.8$), and (e) number of items per dimension ($L = 10, 20$). Each dimension had the same number of items in order to obtain a constant amount of information across dimensions, so that the effect of the number of items would not be confounded with that of the number of dimensions. This setting was also adopted in the literature (Garrido et al. 2011; Svetina 2012). A total of 100 replications in each condition were implemented. For each simulated dataset, the four methods (PA, Hull, DETECT, and SMMC) were adopted, and their performance in detecting the correct dimensionality was compared.

The dependent variable was the accuracy rate, defined as the proportion of times in the 100 replications that the number of dimensions was identified correctly. Additionally, for the SMMC and DETECT methods, it was interesting to know how accurate an item was assigned to its corresponding dimension, given that the number of dimensions had already been identified correctly. The following hit rate was proposed to reflect the accuracy:

$$\frac{\sum_{r=1}^{\text{AR}} \sum_{i=1}^I h_{ri}}{\text{AR} \times I}, \quad (4)$$

where AR was the number of times that the number of dimensions was correctly identified in 100 replications; h_{ri} was a dummy variable and was equal to 1 if an item was correctly assigned to its dimension, and 0 otherwise; and I was the test length. When the hit rate was 1, all items were correctly assigned to their dimensions; when the hit rate was 0, none of the items was correctly assigned.

Parameter configurations were as follows. In the SMMC method, the dimension of the manifolds was set at 1, meaning that each manifold was assumed to be unidimensional; the number of mixtures was set at 2 due to a relatively small number of items; the number of neighbors was set at 2; and the weighting parameter was arbitrarily set at 10. The number of clusters was determined by the eigengap heuristic (Luxburg 2007), which counts the number of times from the first eigenvalue to the last one that the difference in two successive eigenvalues was relatively

smaller than 10^{-6} . As for the PA method, a total of 100 random datasets were generated, and their Pearson correlation matrices were calculated. Factors were extracted using the unweighted least squares with the Promin oblique rotation. Finally, the 95th percentile criterion was used to determine the number of dimensions. The same factor extraction method as in the PA method was adopted for the Hull method, and the comparative fit index was used to describe model-data fit in the Hull method. Regarding the DETECT method, an exploratory approach was adopted, because no prior information about the dimensionality was utilized in the simulations. The minimum number of respondents per score strata was set at 5, meaning that the strata with less than 5 respondents would be removed from the calculation of conditional covariances and then collapsed into adjacent strata. The number of mutations in the genetic algorithm was set at 2 for 10-item tests and 4 for 20-item tests. The maximum number of dimensions for the exploratory search was set at 6. Cross-validation, strongly recommended in small sample sizes or short tests (Monahan et al. 2007), was utilized, in which the respondents were randomly split into two halves to serve as the training and validation subsets. To examine the dimensional structure, the critical values for approximate simple structure index (ASSI) and ratio index (R) were set at 0.25 and 0.36, respectively, which were the default values in poly-DETECT program. When they both were smaller than the critical values, the dataset would be declared as approximate simple structure. The expected conditional covariance was calculated for the affinity matrix.

Only simple structures were considered in this study. In simulating item responses, the discriminate parameter vector α_i was set at $\mathbf{1}$ for the MGPCM; the person parameter vector θ_n was generated from a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance-variance matrix Σ with diagonal elements equal to 1. The correlations among dimensions were all set at a constant. To be consistent with most IRT literature, the person parameters were treated as random effects, but the item parameters as fixed effects, across replications. The intercept parameter δ was set from -2 to 2 with an equal interval between two adjacent items for each dimension. The settings of item and person parameters were generally consistent with simulation and empirical studies in the IRT literature (Muraki 1992). For example, Muraki (1992) uses the range from -1.68 to 1.68 in simulations and obtains a range from -4.47 to 3.37 in an empirical example.

The $\tau_{i1}, \dots, \tau_{iC}$ were drawn from a uniform interval from -2 to 2 with the order of $\tau_{i1} < \tau_{i2} < \dots < \tau_{i(C-1)} < \tau_{iC}$.

We had the following expectations on the simulation results. The PA, Hull, and DETECT methods would be superior to the SMMC method, because the former three methods considered the ordinal nature of categorical data, whereas the latter was a nonparametric method and only consider the continuous nature. The SMMC method might outperform the others as the number of categories is large.

Results

Dichotomous Items

Table 1 shows the accuracy rates and the mean numbers of misspecified dimensions (in parentheses) for two-category items following the one-, two-, and three-dimensional generalized partial credit models. First, consider one-dimensional data. The PA and Hull methods performed almost perfectly across all conditions. The DETECT method did not yield results when the sample size $N = 250$ (DETECT program warned the calculated conditional covariances was inaccurate and did not give results), but yielded good accuracy rates when $N = 1000$ or 2000 . The SMMC method performed satisfactorily only when $N = 1000$ or 2000 , and test length $L = 20$ items (but sometimes, it failed to converge).

Next, consider two- and three-dimensional data. When $N = 250$ and $r = 0$ or 0.4 , the PA and Hull methods outperformed the other two methods in the accuracy rates; when $N = 250$ and $r = 0.8$, the SMMC method performed relatively better. The DETECT method was the best when $N = 1000$ or 2000 . The hit rates were perfect or nearly perfect for the DETECT method and very high for the SMMC method. Generally, it was much more difficult to identify dimensionality when $r = 0.8$ than when $r = 0.4$ or 0 . Fortunately, the DETECT method still yielded a very high accuracy rate when $r = 0.8$, given that $N = 2000$ and $L = 20$.

Polytomous Items

Tables 2 and 3 show the accuracy rates and the mean numbers of misspecified dimensions (in parentheses) for four-category items and six-category items following the one-, two-, and three-dimensional generalized partial credit models, respectively. A comparison of Tables 1 (two-category items), 2 (four-category items), and 3 (six-category items) revealed that the more response categories, the easier the identification of dimensionality. Given the great similarity between Tables 2 and 3, the following discussion focuses on Table 3. First, consider one-dimensional data. The PA method performed perfectly in the identification of the single dimension and outperformed the other three methods. The DETECT method did not yield results when $N = 250$ but performed almost perfectly when $N = 1000$ or 2000 . The Hull method performed perfectly when $L = 10$ but very poorly when $L = 20$ and $N = 1000$ or 2000 . The SMMC method performed satisfactorily when $N = 1000$ or 2000 (but sometimes, it failed to converge).

Next, consider two- and three-dimensional data. When $N = 250$ and $r = 0$ or 0.4 , the PA, Hull, and SMMC methods performed almost perfectly in the identification of the two or three dimensions, whereas the DETECT method did not yield results. When $N = 250$ and $r = 0.8$, the SMMC method was the best. When $N = 1000$ or 2000 , the DETECT method always yielded a perfect accuracy rate. The PA method

Table 1 Accuracy rates, mean numbers of misspecified dimensions (in parentheses), and hit rates (DETECT/SMMC) for two-category items following the one-, two-, and three-dimensional generalized partial credit models

	$N = 250$	$N = 1000$	$N = 2000$
1D/10 items			
PA	1	1	0.99 (2)
Hull	0.99 (2)	1	1
DETECT	NA	0.86 (NA)	0.59 (NA)
SMMC	0.05 (2.2)	0.79 (2)	0.80 (NA)
1D/20 items			
PA	1	0.98 (2)	0.99 (2)
Hull	1	1	1
DETECT	NA	1	1
SMMC	0.06 (2.6)	0.96 (NA)	0.88 (NA)
2D/10 items			
	$N = 250$	$N = 1000$	$N = 2000$
	$r = 0$	$r = 0$	$r = 0$
PA	1	1	1
Hull	0.98 (3)	1	1
DETECT	NA	1	1
SMMC	0.75 (1.5)	0.99 (3)	1
Hit rate	NA/0.99	1/1	1/1
2D/20 items			
	$N = 250$	$N = 1000$	$N = 2000$
	$r = 0$	$r = 0$	$r = 0$
PA	1	1	1
Hull	1	1	1

(continued)

Table 1 (continued)

1D/10 items	N = 250		N = 1000		N = 2000	
	NA	NA	NA	NA	NA	NA
DETECT	0.90 (1.4)	0.54 (1.7)	0.15 (3.9)	0.91 (1)	0.02 (1)	0.13 (1.3)
SMMC	NA/1.00	NA/0.95	NA/0.55	1/1	0.97/0.60	1/1
Hit rate	NA/1.00	NA/0.95	NA/0.55	1/1	0.97/0.60	1/1
3D/10 items	N = 250		N = 1000		N = 2000	
	r = 0	r = 0.4	r = 0.8	r = 0	r = 0.4	r = 0.8
PA	0.99 (4)	0.63 (1.7)	0 (1)	0.99 (4)	0.81 (1.8)	0 (1)
Hull	1	0.38 (1.1)	0 (1)	1	0.99 (1)	0 (1)
DETECT	NA	NA	NA	1	0.01 (1.1)	1
SMMC	0.34 (3.5)	0.30 (3.6)	0.32 (3.8)	0.99 (4)	0.67 (1.9)	0 (1.0)
Hit rate	NA/0.87	NA/0.61	NA/0.38	1/1	1.00/1.00	0.78/NA
3D/20 items	N = 250		N = 1000		N = 2000	
	r = 0	r = 0.4	r = 0.8	r = 0	r = 0.4	r = 0.8
PA	1	1	0 (1)	1	0 (1.1)	0 (1.1)
Hull	1	0.74 (1)	0 (1)	1	0 (1)	0 (1)
DETECT	NA	NA	NA	1	0.04 (1)	1
SMMC	0.45 (2.2)	0.19 (2.8)	0.19 (4.8)	1	0.75 (2)	0 (1)
Hit rate	NA/0.99	NA/0.70	NA/0.42	1/1	1/1	0.90/NA

Note NA not available; hit rates are for the DETECT and SMMC methods

Table 2 Accuracy rates, mean numbers of misspecified dimensions (in parentheses), and hit rates (DETECT/SMIMC) for four-category items following the one-, two-, and three-dimensional generalized partial credit models

	<i>N</i> = 250	<i>N</i> = 1000	<i>N</i> = 2000
1D/10 items			
PA	1	1	1
Hull	0.92 (2)	1	1
DETECT	NA	0.96 (2)	0.97 (2)
SMIMC	0.17 (2.1)	0.74 (2)	0.74 (NA)
1D/20 items	250	1000	2000
PA	1	1	1
Hull	0.92 (2)	0.39 (2)	0.05 (2)
DETECT	NA	0.99 (NA)	1
SMIMC	0.25 (2.3)	0.90 (2)	0.86 (2)
2D/10 items	<i>N</i> = 250	<i>N</i> = 1000	<i>N</i> = 2000
	<i>r</i> = 0	<i>r</i> = 0	<i>r</i> = 0
PA	1	1	1
Hull	0.99 (3)	0 (1)	0 (1)
DETECT	NA	NA	1
SMIMC	0.97 (3)	0.55 (2.3)	0.98 (NA)
Hit rate	NA/1	NA/0.72	1/1
2D/20 items	<i>N</i> = 250	<i>N</i> = 1000	<i>N</i> = 2000
	<i>r</i> = 0	<i>r</i> = 0	<i>r</i> = 0
PA	1	1	1
Hull	0.99 (3)	0 (1)	0 (1.3)
DETECT	NA	NA	1
SMIMC	0.97 (3)	0.93 (2.6)	0.92 (1)
Hit rate	NA/1	NA/0.72	1/1
2D/20 items	<i>N</i> = 250	<i>N</i> = 1000	<i>N</i> = 2000
	<i>r</i> = 0	<i>r</i> = 0	<i>r</i> = 0
PA	1	1	1
Hull	0.99 (3)	0 (1.1)	0 (1.5)
DETECT	NA	NA	1
SMIMC	0.97 (3)	0.55 (2.3)	0.92 (1)
Hit rate	NA/1	NA/0.72	1/1
2D/20 items	<i>N</i> = 250	<i>N</i> = 1000	<i>N</i> = 2000
	<i>r</i> = 0	<i>r</i> = 0	<i>r</i> = 0
PA	1	1	1
Hull	0.99 (3)	0 (1.1)	0 (1.5)
DETECT	NA	NA	1
SMIMC	0.97 (3)	0.55 (2.3)	0.92 (1)
Hit rate	NA/1	NA/0.72	1/1

(continued)

Table 2 (continued)

	N = 250		N = 1000		N = 2000	
1D/10 items	NA	NA	NA	1	1	1
DETECT	0.99 (3)	0.99 (4)	0.37 (1.5)	0.99 (NA)	0.84 (1)	0.98 (NA)
SMMC	NA/1	NA/1.00	NA/0.77	1/1	1/1	1/1
Hit rate						
3D/10 items	N = 250		N = 1000		N = 2000	
	r = 0	r = 0.4	r = 0.8	r = 0	r = 0.8	r = 0.8
PA	1	1	0 (1)	1	0 (1.0)	1
Hull	1	0.99 (1)	0 (1)	1	0 (1.1)	1
DETECT	NA	NA	NA	1	1	1
SMMC	0.96 (4)	0.84 (4.3)	0.27 (2.7)	0.98 (4)	0.41 (1.7)	0.98 (NA)
Hit rate	NA/1	NA/1.00	NA/0.57	1/1	1.00/1	1/1
3D/20 items	N = 250		N = 1000		N = 2000	
	r = 0	r = 0.4	r = 0.8	r = 0	r = 0.8	r = 0.8
PA	1	1	0.48(3.1)	1	0.05(4.0)	0.99(4)
Hull	0.99 (4)	1	0 (1)	1	0 (1.1)	0 (1.2)
DETECT	NA	NA	NA	1	1	1
SMMC	0.98 (4)	0.98 (3)	0.29 (2.6)	1	0.61 (1.9)	0.98 (4)
Hit rate	NA/1	NA/1	NA/0.51	1/1	1/1	1.00/1

Note NA not available; hit rates are for the DETECT and SMMC methods

Table 3 (continued)

1D/10 items	N = 250		N = 1000		N = 2000				
	NA	NA	NA	NA	1	1			
DETECT	0.97 (3)	0.97 (3.3)	0.40 (1.3)	0.96 (3)	0.94 (3)	0.92 (2.6)	0.98 (3)	0.95 (3)	0.95 (NA)
Hit Rate	NA/1	NA/1	NA/0.96	1/1	1/1	1/1	1/1	1/1	1/1
3D/10 items	N = 250		N = 1000		N = 2000				
	r = 0	r = 0.4	r = 0.8	r = 0	r = 0.4	r = 0.8	r = 0	r = 0.4	r = 0.8
PA	1	1	0 (1.0)	1	1	0 (1.0)	1	1	0 (1)
Hull	1	1	0 (1)	1	1	0 (1.3)	1	1	0 (1.6)
DETECT	NA	NA	NA	1	1	1	1	1	1
SMMC	0.98 (4)	0.97 (4)	0.31 (3.1)	0.95 (4)	0.96 (4)	0.92 (2.8)	1	0.93 (4)	0.95 (NA)
Hit rate	NA/1	NA/1	NA/0.72	1/1	1/1	1/1	1/1	1/1	1/1
3D/20 items	N = 250		N = 1000		N = 2000				
	r = 0	r = 0.4	r = 0.8	r = 0	r = 0.4	r = 0.8	r = 0	r = 0.4	r = 0.8
PA	1	0.99 (4)	0.23 (4.0)	1	1	0 (4.0)	1	1	0 (4.0)
Hull	1	1	0 (1.1)	0.97 (6)	1	0 (1.7)	0.26 (6)	0.94 (6)	0 (1.8)
DETECT	NA	NA	NA	1	1	1	1	1	1
SMMC	0.97 (4)	0.92 (4)	0.32 (3.2)	0.92 (4)	0.91 (4)	0.92 (3.6)	0.92 (4)	0.93 (4)	0.91 (4)
Hit rate	NA/1	NA/1	NA/0.81	1/1	1/1	1/1	1/1	1/1	1/1

Note NA not available; hit rates are for the DETECT and SMMC methods

always yielded a perfect accuracy rate when $r = 0$ or 0.4 , but it performed very poorly when $r = 0.8$. The Hull method performed poorly when $L = 20$ or $r = 0.8$. The SMMC performed almost perfectly across all conditions. The hit rates were always perfect for the DETECT method, except when the sample size was 250, and often perfect for the SMMC method. In summary, the following recommendations appear applicable:

1. The PA method is recommended when sample sizes are small, and the number of dimension is one or the correlations among dimensions are low.
2. The DETECT method is recommended when sample sizes are large.
3. The SMMC method can be a supplement to the PA and DETECT methods when the number of categories is large.

Conclusion and Discussion

The SMMC method makes no assumption on the responding process, seeming to be a promising alternative for dimensionality assessment. Although the SMMC performs worse than other methods in general, it can serve as a supplement to the PA and DETECT methods.

Future studies can aim at evaluating these four methods and other methods under a more comprehensive design. For example, multidimensional scaling and bootstrap generalization are promising methods of dimensionality assessment (Finch and Monahan 2008; Meara et al. 2000). In this study, only simple structures were investigated. The dimensionality assessment of complex structures, in which an item may measure more than one dimension and the dimensions can be compensatory or noncompensatory (Embretson 1997; Symпсо 1978; Whitely 1980), is of great importance and left for future studies. The PA method used in this study adopted the Pearson correlation in order to be consistent with the literature (Timmerman and Lorenzo-Seva 2011; Weng and Cheng 2005). Actually, the polychoric correlation can be used for ordinal data (Cho et al. 2009; Timmerman and Lorenzo-Seva 2011).

References

- Adams, R. J., Wilson, M., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1–23.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores (Chaps. 17–20)*. Reading, MA: Addison-Wesley.
- Cho, S.-J., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement, 69*(5), 748–759.
- Embretson, S. E. (1997). Multicomponent response models. *Handbook of modern item response theory* (pp. 305–321), Springer.

- Finch, H., & Monahan, P. (2008). A bootstrap generalization of modified parallel analysis for IRT dimensionality assessment. *Applied Measurement in Education, 21*(2), 119–140.
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's minimum average partial factor retention method with categorical variables. *Educational and Psychological Measurement, 71*(3), 551–570.
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). Factor: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, Instruments, & Computers, 38*, 88–91.
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research, 46*(2), 340–364.
- Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing, 17*(4), 395–416.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.
- Meara, K., Robin, F., & Sireci, S. G. (2000). Using multidimensional scaling to assess the dimensionality of dichotomous item data. *Multivariate Behavioral Research, 35*(2), 229–259.
- Monahan, P. O., Stump, T. E., Finch, H., & Hambleton, R. K. (2007). Bias of exploratory and cross-validated DETECT index under unidimensionality. *Applied Psychological Measurement, 31*(6), 483–503.
- Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159–176.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danmarks Paedagogiske Institut, 1960. Chicago: University of Chicago Press, 1980.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer-Verlag.
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(8), 888–905.
- Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L., & Zhang, Jinming. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement, 20*(4), 331–354.
- Svetina, D. (2012). Assessing dimensionality of noncompensatory multidimensional item response theory with complex structures. *Educational and Psychological Measurement*.
- Svetina, D., & Levy, R. (2014). A framework for dimensionality assessment for multidimensional item response models. *Educational Assessment, 19*(1), 35–57.
- Symonson, J. B. (1978). *A model for testing with multidimensional items*. Paper presented at the Proceedings of the 1977 computerized adaptive testing conference.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209–220.
- Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61*(3), 611–622.
- 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*(1), 50–61.
- Wang, Y., Jiang, Y., Wu, Y., & Zhou, Z.-H. (2011). Spectral clustering on multiple manifolds. *IEEE Transactions on Neural Networks, 22*(7), 1149–1161.
- Weng, L.-J., & Cheng, C.-P. (2005). Parallel analysis with unidimensional binary data. *Educational and Psychological Measurement, 65*(5), 697–716.
- Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika, 45*(4), 479–494.
- Wilson, D. T., Wood, R., & Gibbons, R. D. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis*. SSI, Scientific Software International.
- Yao, L., & Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: an application to mixed-format tests. *Applied Psychological Measurement, 30*(6), 469–492.
- Zhang, J. (2007). Conditional covariance theory and detect for polytomous items. *Psychometrika, 72*(1), 69–91.
- Zhang, J., & Stout, W. (1999). The theoretical detect index of dimensionality and its application to approximate simple structure. *Psychometrika, 64*(2), 213–249.