Original Article

Check for updates

Partitioning Dichotomous Items Using Mokken Scale Analysis, Exploratory Graph Analysis and Parallel Analysis: A Monte Carlo Simulation

Gomaa Said Mohamed Abdelhamid [1,2] ⓘ , María Dolores Hidalgo [3] ⓘ ,

Brian F. French [4] ⓘ , Juana Gómez-Benito [5,6] ⓘ

[1] *Department of Educational Psychology, Faculty of Education, Fayoum University, Fayoum, Egypt.* [2] *Department of Psychology, College of Education, Sultan Qaboos University, Muscat, Oman.* [3] *Department of Basic Psychology and Methodology, Faculty of Psychology, University of Murcia, Murcia, Spain.* [4] *Department of Kinesiology and Educational Psychology, Washington State University, Pullman, WA, USA.* [5] *Department of Social Psychology and Quantitative Psychology, Faculty of Psychology, University of Barcelona, Barcelona, Spain.* [6] *Group on Measurement Invariance and Analysis of Change (GEIMAC), Institute of Neuroscience, University of Barcelona, Barcelona, Spain.*

Abstract

Estimating the number of latent factors underlying a set of dichotomous items is a major challenge in social and behavioral research. Mokken scale analysis (MSA) and exploratory graph analysis (EGA) are approaches for partitioning measures consisting of dichotomous items. In this study we perform simulation-based comparisons of two EGA methods (EGA with graphical least absolute shrinkage and selector operator; EGAtmfg with triangulated maximally filtered graph algorithm), two MSA methods (AISP: automated item selection procedure; GA: genetic algorithm), and two widely used factor analytic techniques (parallel analysis with principal component analysis (PApc) and parallel analysis with principal axis factoring (PApaf)) for partitioning dichotomous items. Performance of the six methods differed significantly according to the data structure. AISP and PApc had highest accuracy and lowest bias for unidimensional structures. Moreover, AISP demonstrated the lowest rate of misclassification of items. Regarding multidimensional structures, EGA with GLASSO estimation and PApaf yielded highest accuracy and lowest bias, followed by EGAtmfg. In addition, both EGA techniques exhibited the lowest rate of misclassification of items to factors. In summary, EGA and EGAtmfg showed comparable performance to the highly accurate

traditional method, parallel analysis. These findings offer guidance on selecting methods for dimensionality analysis with dichotomous indicators to optimize accuracy in factor identification.

Keywords

exploratory graph analysis, Mokken scale analysis, latent factors, community detection, dichotomous data, dimensionality, network psychometrics, parallel analysis

Dimensionality analysis is an important step in scale or questionnaire development. The determination of the theoretically consistent number of latent factors to retain can be a major challenge for researchers (Lorenzo-Seva & Ferrando, 2023), especially with dichotomous indicators or items (Lubbe, 2019). Many standard estimation techniques in dimensionality assessment, such as normal theory maximum likelihood (ML), assume that variables are continuous and normally distributed (Rhemtulla et al., 2012), which is violated with dichotomous variables. Therefore, relying on standard or traditional methods such as parallel analysis (PA; Horn, 1965) for estimating the dimensionality of the latent structure with dichotomous indicators with accuracy has become a central issue in many simulation studies (Antino et al., 2018; Ayanwale et al., 2020; Finch, 2020; Lubbe, 2019), as well as software selection (e.g., Chen & Zhang, 2021; Lee & Cham, 2024; Svetina & Levy 2012) for such work. Standard factor analytic approaches applied to categorial, and binary data tend to identify spurious factors in the presence of noncentered distributions. A main issue is that the magnitude of the correlation between the binary items depends on the: (a) magnitude of the true relationship, and (b) the response rate of alternative item responses. The difference in prevalence can produce subgroups of items that are spuriously related and lead to detection of spurious factors (Bernstein & Teng 1989). Even with adjustment techniques (e.g., DWLS, polychoric correlations) to avoid spurious factors, nonconvergence issues persist, especially with small sample sizes and asymmetrical item response distributions (e.g., Flora & Curran, 2004). The accuracy of techniques to identify the number of factors, such as parallel analysis, are also influenced by the distributional characteristics. Adjustments for these data conditions tend to improve accuracy, depending on data conditions (Lubbe, 2019), but are not perfect.

Several alternative methodologies for examining the latent structures of dichotomous item scales have been described. Among these, exploratory graph analysis (EGA) has been proposed as an accurate method for assessing the correct number of factors (Golino et al., 2020; Golino & Epskamp, 2017). EGA is a technique based on the network psychometrics approach (Epskamp et al., 2017) and first restricted to multivariate normal distributions. So, first estimates a network of variables (i.e., nodes) using the Gaussian graphical model (GGM; Lauritzen 1996) and applies the graphical least absolute shrinkage and selector operator (GLASSO; Friedman et al. 2008). Second, the *walktrap* algorithm (Pons & Latapy, 2006) is utilized to determine the communities or clusters of variables via the connections (i.e., edges/links) between the nodes (i.e., test items) of the estimated network. These edges represent partial correlation coefficients, where

the edges depict the strength of the connection between items (Epskamp et al., 2018). Notably, these communities are theoretically comparable to latent factors (Golino & Epskamp, 2017).

Golino et al. (2020) proposed an alternative network estimation method for EGA, the Triangulated Maximally Filtered Graph (TMFG) algorithm (Massara et al., 2016) as an alternative network estimation method for EGA. Specifically, they proposed substituting GGM with the TMFG algorithm in order to overcome drawbacks of using EGA with the former (Golino et al., 2020). The advantages of the TMFG algorithm include: (a) application with normal and non-normal data, (b) not constrained to partial correlation measures, and (c) stable performance across different sample sizes. Thus, EGA method with the TMFG algorithm is not restricted to multivariate normal distributions and partial correlation measures. Any association measure can be used, allowing its application with binary or polytomous variables.

Figure 1 depicts a visualization comparison of a simulated five-factor model with interfactor correlations of zero (e.g., orthogonal factors; Panel A) and .6 (multiple correlated factors; Panel B) and a sample size of 1000 using EGA with GLASSO estimation and TMFG. This figure shows a network of nodes connected by strong or weak edges based on their correlation strength. Nodes with higher correlations are closer together, while those with lower correlations are further apart. The figure is created using the EGA approach, which helps identify variable groupings by clustering variables with higher correlations. The nodes are color-coded to represent different factors. By using regularized partial correlations, we can obtain a clearer structure with five groups of variables in the high correlation structure. Additionally, the regularized partial correlations show stronger connections within clusters than between clusters for the high correlation structure (on the right), which helps depict the true simulated five-factor structure, even when the true correlation between factors is high.

Item response theory (IRT) has been used in educational and psychological measurement (Reise & Waller, 2009; Thomas, 2019). Mokken scale analysis (MSA; Mokken 1971) is IRT approach to constructing questionnaires and tests in various fields (Sijtsma et al., 2008; Wind & Wang, 2023). van der Ark et al., (2008) classified MSA into two basic steps. The first step evaluates a test with several ordered items using the monotone homogeneity model in either a confirmatory or an exploratory approach. The confirmatory approach is applied in the case of items that are assumed to form a single scale, whereas the exploratory approach is used to determine whether a test contains one or more scales (van der Ark et al., 2008). The first step of MSA also involves an exploratory assessment to partition an item pool into different clusters. The second step of MSA is concerned with the investigation of the psychometric properties of the scales identified in the first part.

The overall objective of MSA is to select from a given pool of items, where clustered items range from weak to strong discrimination, as many sufficiently-discriminating

Figure 1

A Visualization Comparison of a Simulated Five-Factor Model With Interfactor Correlations of Zero (Left) and .6 (Right) and a Sample Size of 1000 Using EGA With Glasso Estimation and With TMFG
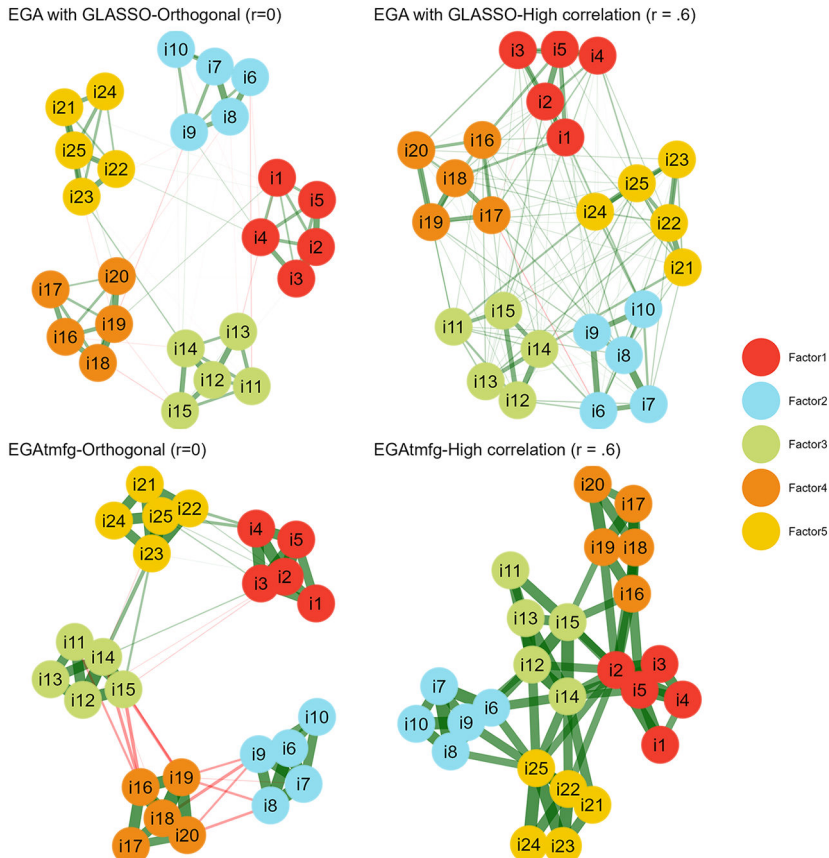
Note. Nodes correspond to variables, edges correspond to correlations, and the node colors correspond to the simulated factors. This figure is extracted from our simulation data.

items as possible in each cluster (Mokken, 1971; Straat et al., 2013). Typically, the researchers define what they consider to be "sufficient" discrimination (Straat et al., 2013:77), as sufficient discrimination may differ on the construct measured and purpose of scores (e.g., norm vs. criterion decisions). To this end, Mokken (1971) suggested an *automated item selection procedure* (AISP) to partition an item set into a unidimensional cluster or multiple clusters depending on the data, where each cluster assesses the same latent trait and satisfies particular scaling criteria (Mokken, 1971, pp. 189–190). Each

cluster is referred to as the Mokken scale that involves sufficiently-discriminating items. Straat et al. (2013) proposed the *genetic algorithm* as an alternative to the AISP for optimal item partitioning into Mokken scales. The aim of both Mokken's original AISP and the later GA method is to find the largest Mokken scale possible with sufficiently-discriminating items, and to form from the remaining items a second largest Mokken scale possible with sufficiently-discriminating items. If the data allow, the process may continue to form the largest additional scales (i.e., the third, fourth,...etc.) and perhaps detect one or more unscalable items (Mokken, 1971; Straat et al., 2013).

Notably, although traditional factor-analysis methods are widely used, many of these methods have limitations with categorical data (Golino et al., 2020). For instance, Garrido et al. (2013) noted that parallel analysis is highly sensitive to various factors, including sample size, factor loadings, number of variables per factor, and factor correlations. Additionally, factor analytic techniques present challenges beyond dimension estimation, such as subjective interpretation of factor loadings and rotation of the loadings matrix (Sass & Schmitt, 2010).

Therefore, many authors have suggested using the exploratory MSA approach and network psychometrics approach to investigate the internal structure of dichotomous item scales (Abdelhamid et al., 2020; Mokken, 1971; Nuño et al., 2022). Furthermore, MSA techniques and EGA methods used to extract data dimensions have shown promise compared to traditional factor analysis (parallel analysis) methods, mainly in sample sizes of 500 or higher and in conditions with number of Factors 4 and 6 (Antino et al., 2018; Golino et al., 2020; Haslbeck & van Bork, 2024; Straat et al., 2013; van Abswoude et al., 2004).

The Current Study

This investigation focused on methods that can verify the internal structure of scales containing dichotomous items. Importantly, these methods can be applied to a wide range of research contexts in psychology, and more broadly the social and behavioral sciences, where an alternative to factor analysis is warranted due to the data at hand. This study assessed the performance of EGA (EGA with GLASSO and EGAtmfg), MSA (AISP and GA) techniques and parallel analysis with principal component analysis (PApc) and parallel analysis with principal axis factoring (PApaf) with respect to the proportion of the number of factors underlying the latent structure, parameter bias, root mean squared deviation, and misclassification rate. To this end, we examined the effect of sample size, factor intercorrelation, and the number of indicators per factor on the accuracy of EGA with GLASSO, EGAtmfg, GA, AISP, PApaf and PApc using Monte Carlo simulation. Because our study is concerned with scales or tests, the indicators of the latent variable are called items.

The current study differs in several respects from earlier research into the performance of EGA and MSA methods. Previous studies on the performance of EGA assessed

PsychOpen GOLD

only the accuracy of the number of factors, regardless of whether items were correctly classified on their corresponding factor (Golino et al., 2020; Haslbeck & Van Bork, 2024). Here, by contrast, we assess both the accuracy of the number of factors and the classification of items to the correct factor. The latter is of interest since it is possible that number of factors is retrieved correctly but factors may contain items that are not classified correctly, which negatively affects the accuracy of these methods. Their accuracy must therefore be examined with respect to two criteria: correct item classification on the true factor and correct estimate of the number of factors. This corresponds to the main aim of the current study. Regarding the number of latent variables (e.g., factors), previous simulation studies concerned with Mokken methods (Straat et al., 2013) were limited to two factors. Furthermore, previous studies of EGA are limited to just one, two, three, or four factors.

The current study therefore seeks to evaluate the performance of these methods across structures that are widely used in psychological and sociological measures (e.g., one, two, four, and five factors). Finally, the number of replications per condition in previous studies was only 100 or 500 (Golino et al., 2020; Straat et al., 2013), whereas in the current study the number of replications is expanded to 1000 so as to achieve accuracy in the conclusions drawn.

Theoretical expectations regarding the performance of EGA methods versus MSA and parallel analysis methods depend on various assumptions and properties inherent to both methodologies. We anticipate that the EGA techniques will perform better in terms of accuracy of factor identification compared to Mokken techniques, especially in the condition of scales that contain highly correlated factors. This is because EGA methods are designed to identify sparse and low-rank substructures in high-dimensional data, which can be helpful for detecting highly correlated factors. EGA methods use regularization techniques (L1 or LASSO), its objective is not to reduce the variance of the individual estimates, but rather to improve interpretability and select a more parsimonious model. Despite the high variance in highly correlated situations, L1 regularization helps to select a subset of relevant variables by forcing some coefficients to be exactly zero. This selection process can be useful in identifying the number of really important factors, even when the variables are highly correlated, as it tends to keep only one of the factors correlated, thus reducing the multicollinearity problem. On the other hand, MSA methods may struggle to identify these correlations since they tend to focus on finding groups of variables with similar distributions or patterns.

Additionally, we expect that the EGA techniques will perform better than MSA techniques in terms of misclassification rate. This is because EGA methods usually use regularization techniques like the least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996) to estimate coefficients, which reduce parameter bias and improve model performance. Conversely, MSA methods may suffer from overfitting or underfitting since they focus on grouping variables based on their distribution or pattern

rather than using regularization techniques. Furthermore, we anticipate accuracy and bias to degrade in conditions with smaller sample sizes, and larger number of factors and indicators. Finally, we expect that EGA methods will perform comparably to the highly accurate traditional method of parallel analysis (Golino et al., 2020; Golino & Epskamp, 2017).

It is important to acknowledge that the expectations presented are based on certain assumptions and may not always be applicable in applied contexts. The effectiveness of these methods can be affected by several factors such as sample size, factor intercorrelation, and the number of indicators per factor. Therefore, it is crucial to assess the performance of each method through appropriate simulation studies before drawing conclusions about their respective advantages.

Method

Simulation Model

Dichotomous data were generated using a multidimensional two-parameter logistic IRT (M2PL) model (Reckase, 2009). The multidimensional 2PL formula is written as:

$$P_i\left(\theta_j\right) = \frac{e^{D(a_i\theta'_j + b_i)}}{1 + e^{D(a_i\theta'_j + b_i)}} \tag{1}$$

$$\text{where } a_i\theta'_j + d_i = a_{i1}\theta_{j2} + a_{i2}\theta_{j2} + \dots + a_{im}\theta_{jm} + d_i = \sum_{l=1}^{m} a_{il}\theta_{jl} + b_i \tag{2}$$

where θ'_j represents m latent factors ($\theta'_j = \theta_{j1}, \dots, \theta_{jm}$) for Subject j, a_i indicates item slopes (discrimination parameters: $a_i = a_1, \dots, a_m$) for item i, b_i represents the item intercept term (difficulty) for Item i, and D indicates a scaling adjustment (usually 1.702).

Equation (1) can be written as:

$$P_i\left(\theta_j\right) = \frac{1}{1 + e^{-D(a_i\theta'_j + b_i)}} \tag{3}$$

Notably, Multidimensional Item Response Theory models are frequently employed to develop and investigate the psychometric properties of measures in educational and psychological assessments. These models are used for analyzing dichotomous data in questionnaires related to general health, education, and personality (Ackerman et al., 2003; Ayanwale et al., 2024; Chernyshenko et al., 2001; Immekus et al., 2019; Li et al., 2012).

Procedures for Examining Dimensionality

GLASSO

The default EGA (using the GGM model) is based on the GLASSO estimation, which utilizes the LASSO technique (Tibshirani, 1996). This method aims to achieve a sparse network model that can be built at various levels, from a fully connected network to a totally disconnected network.

TMFG Algorithm

The TMFG algorithm is a filtering method, which has been proposed as a way of detecting communities/clusters in a network (Massara et al., 2016). This method constructs a triangulation that maximizes a score function (e.g., the sum of the edge weights) related to the quantity of information preserved by the network. This method filters a network to retain a total of 3n−6 edges (n indicates the number of nodes) from the estimated network through the planarity constraint (Christensen et al., 2018). This method first constructs a sub-network from the original network with three-node cliques (a triangle with connected nodes) based on zero-order correlations. Second, a node is added to this sub-network using a score function that maximizes the sum weights of the three connecting edges. This sub-network will be a four-node clique (called tetrahedron, with the highest overall total weights score), insofar as a node is added to the three-node cliques (Christensen et al., 2018).

Golino et al. (2020) proposed two EGA techniques: an extension of EGA by expanding the correlation matrix to have orthogonal correlations ("expand") to deal correctly with unidimensional structures, and EGA using the TMFG algorithm (EGAtmfg). They utilized the EGA approach to "expand" the empirical correlation matrix by adding four variables that are highly correlated with each other (r = .50; roughly equivalent to factor loadings of .70) and completely uncorrelated (0.00) with all other empirical variables. They compared these methods with five traditional methods and found that EGA, EGAtmfg, and parallel analysis performed well. Furthermore, EGA with GLASSO estimation was more accurate in terms of percentage of correct numbers of factors recovery than the TMFG estimation, suggesting EGA with GLASSO may be the preferred method.

Furthermore, Christensen et al. (2024) adjusted the EGA approach by applying the Leading Eigenvalue ("LE") community detection algorithm (Newman, 2006) to the correlation matrix, This adjustment proved particularly effective in enhancing EGA's performance in handling unidimensional structures. The LE algorithm utilizes the first eigenvector of the modularity matrix to determine the number of communities. It accomplishes this by iteratively dividing up the network into two communities until further modularity improvements are no longer attainable.

PsychOpen GOLD

The Automated Item Selection Procedure (AISP)

The automated item selection procedure (AISP), known as a bottom-up item selection method, forms a cluster of items in several steps (Sijtsma & Molenaar 2002; Wismeijer et al., 2008). First, the item pair with the greatest scalability coefficient value H_{jk1}, which is much greater than 0 and surpasses the lower bound c, is chosen from any $\frac{1}{2}J(J-1)$ item pairs. Let item pair (j, k_1) be the starting pair for constructing a Mokken Scale 1. A third item, k_2, is then chosen for Mokken Scale 1 from the remaining items J-2, with the following conditions being fulfilled: (i) this third item correlates positively with the previously chosen items j and k_1, (ii) it has a scalability value of H_{jk2} and H_{k1k2} is considerably above 0 and surpasses the lower bound c, and (iii) the chosen items j, k_1 and k_2 provide the maximum scalability H for Mokken Scale 1, compared with the remaining items. This process is then replicated with a fourth item, a fifth item, etc. The AISP will stop on a selected item for Mokken Scale 1 if there are no remaining items that achieve the conditions. If the data allow (i.e., they are multidimensional), the AISP forms Scale 2, Scale 3, Scale 4, and so on from the remaining items. Note that some items may be unscalable. More detailed information about the Mokken Scale Analysis (MSA) and scalability coefficients is available online in the supplementary materials by Abdelhamid et al. (2024a).

Genetic Algorithm Procedure

Straat et al. (2013) suggested the GA procedure based on a genetic algorithm for partitioning test items into Mokken scales. The GA attempts to select the best partitioning that meets Mokken's objective among all simultaneously possible partitionings of the item pool. The GA method produces an initial population that involves random partitionings and then examines each partitioning according to Mokken's goal. The initial population with random partitioning is then replaced by the second random population. The probability of choosing the same partitioning that best achieved Mokken's goal in the first population increases in the next random population of partitioning. The GA then assesses all partitioning in the second population and generates a third population using the same procedures as for the second population. It identifies the best partitioning based on Mokken's goal in each population of random partitioning up to the latest population. Finally, the GA reports the best partitioning, which is invariant across almost all populations (Straat et al. 2013). In recent years, the GA has been used to select Mokken scales (Abdelhamid et al., 2020; Ahmadi et al., 2016).

Parallel Analysis Comparison

We utilized two parallel analysis algorithms (PApaf and PApc). We chose these algorithms due to their wide-ranging evaluation in the literature (for example, Garrido et al., 2013) and their comparable performance with EGA in a previous simulation study (Golino et al., 2020). In simpler terms, parallel analysis involves creating a large number

PsychOpen GOLD

of replicate datasets by randomly selecting values from each variable in the original dataset (Horn, 1965). The suggested factor option is the number of factors (PApaf) or components (PApc) whose eigenvalues in the original dataset are greater than the mean of the resampled datasets. We conducted parallel analysis based on the tetrachoric correlation coefficient due to the nature of data, which is dichotomous.

Design Conditions

We included a wide range of conditions that frequently appear in empirical studies using scales or questionnaires, as follows:

1. *Sample size (three levels).* Sample size was set to 250, 1000 or 5000 for each data generation model, thus covering the minimum of several observations required for structure analysis (Comrey & Lee, 1992; Golino et al., 2020; Savalei & Rhemtulla, 2013; Straat et al., 2014). Thus, sample sizes of 250 and 1000 can be considered as sample size medium and large while a sample of 5000 observations allows for the evaluation of the dimensionality methods in conditions that can approximate their population performance (Golino et al., 2020).
2. *Number of factors (four levels).* The number of factors considered was one, two, four, and five, which reflects the dimensionality of most measuring instruments (Garrido et al., 2016; Henson & Roberts, 2006).
3. *Factor correlations (three effect size levels).* The correlation between factors was .0, .30, or .60, representing multidimensional scales whose factors are orthogonal and multidimensional scales whose factors are moderately and highly correlated (Cohen, 1988), respectively.
4. *Number of items per latent factor (eight levels for unidimensional conditions, two levels for multidimensional conditions).* The number of items for unidimensional conditions was set to 5, 8, 10, 16, 20, 25, 32, and 40 items, covering the range from short to long scales. For multidimensional conditions (i.e., two, four or five underlying factors), the number of items per factor was set to five or eight, above the required minimum of three items for identification of a latent factor (Velicer et al., 2000; Widaman, 1993).

Two characteristics were fixed in the simulation design. First, the values of item discrimination parameters ranged from .8 to 1.4, with the mean discrimination fixed at 1. Second, to ensure a simple structure, and considering two factors as an example, discrimination parameters of the first half of the items (first-factor) had a mean $a_{j1} = 1$ and $a_{j2} = 0$, whereas for the second half of the items (second-factor) the mean was $a_{j1} = 0$ and $a_{j2} = 1$. In addition, the d parameter was drawn from a uniform distribution on the interval [−2, 2].

PsychOpen GOLD

Data Generation

The total number of unidimensional conditions for data generation was 24, corresponding to 3 sample sizes × 1 number of factors × 8 number of items per factor. The total number of multidimensional conditions for data generation was 54 conditions, corresponding to 3 sample sizes × 3 number of factors × 2 number of items per factor × 3 factor correlations. For each condition, 1000 independent random data sets were generated in R (R Core Team, 2021). A total of 78 conditions combinations were studied. For each of these 78 conditions combinations, 1000 replicates were simulated.

Data Analysis

All data sets were analyzed using R packages and four methods: (1) the package 'EGAnet' (Golino & Christensen, 2020) was used to conduct all analyses for EGA with GLASSO and EGAtmfg, (2) the package 'Mokken' (van der Ark, 2012) was used to analyze all simulated data sets when using the two MSA methods (AISP and GA) with the default lower bound (c = .3); the lower bound c indicates the minimum level of discrimination (H_j) that items must meet to be included in the Mokken scale. In other words, items with a discrimination value below this threshold will not be included in the scale, (3) the package "psych" (Revelle, 2024) was used to conduct the parallel analysis using the fa.parallel function. An example of R syntax for conducting EGA, MSA methods, and Parallel Analysis is available online in Abdelhamid et al. (2024c) to facilitate the implementation of these methods.

Four complementary criteria were applied to assess the performance of methods:

1. The proportion of correctly estimated numbers of factors (pc) across all simulation replications was calculated for each method by:

$$pc = \frac{\sum T}{N} \text{for T} = \begin{cases} 1 & \text{if} & \widehat{F}_i = & F_{True} \\ 0 & \text{if} & \widehat{F}_i \neq & F_{True} \end{cases} \quad (4)$$

where $\widehat{F}_i$ is the number of factors identified by each method for replication i, F_{True} is true dimensionality, and N represents the number of sample data generated for each condition (1000). True dimensionality is the extent to which each technique retrieves the true number of factors in the simulated data sets. The pc criterion ranges from 0 (representing complete inaccuracy) to 1 (representing perfect accuracy).

2. The mean bias error (MBE) was estimated by:

$$MBE = \frac{\sum_i^N \left(\widehat{F}_{i=1} - F_{True} \right)}{N} \quad (5)$$

The MBE indicates the average deviation from the correct number of factors where a negative value is considered as under-factoring, positive values indicate over-

factoring and 0 indicates a total lack of bias. Mean bias error, instead of a relative bias measure, is used as a measure of accuracy because it provides a direct estimate of the amount of error in the parameter estimates.

3. The root mean squared deviation (RMSD) of the number of factors is reported for each method. The RMSD is expressed by:

$$RMSD = \sqrt{\frac{\sum_i^N \left(\widehat{F}_{i=1} - F_{True}\right)^2}{N}} \tag{6}$$

The RMSD is an indicator of the average deviation from the true factor number, and it also provides information about the dispersion of procedure estimates. A large value of the RMSD indicates greater variance in the number of factors reported by a method and the value of 0 represents perfect estimation accuracy.

4. Misclassification rate (%), which refers to the percentage of incorrectly classified items on a factor. This can be estimated as:

$$\text{Misclassification rate } (MR; \%) = \frac{\sum F}{\text{Total items}},$$
$$F = \begin{cases} 0 & \textit{correctly classified item} \\ 1 & \textit{incorrectly classified item} \end{cases} \tag{7}$$

The best method is the one that achieves the minimum rate of item misclassification and the value of 0 reflects perfect estimation accuracy.

Results

The performance of the six methods (EGA with GLASSO, EGAtmfg, AISP, GA, PApaf, and PApc) was evaluated separately for designs with: (i) unidimensional conditions, (ii) multiple orthogonal factors, and (iii) multiple correlated factors. The results were presented both graphically and descriptively. For each design, the main results are summarized as follows: (1) the best-performing method generally was determined based on *pc*, MR, MBE, and RMSD, followed by (2) a discussion of which method achieves better results than others under different conditions.

Unidimensional Conditions

The EGAnet package offers two options to the unidimensional problem, including "expand" and "leading eigenvalue" (LE). Therefore, the current study tested both options for the two methods of EGA (e.g., EGA with GLASSO under expand option [EGA_expand], EGA with GLASSO under LE option [EGA_LE], EGAtmfg under expand option [EGAtmfg_expand], EGAtmfg under LE option [EGAtmfg_LE]). Figure 2 and Figure 3 summarize the average performance of the EGA, Mokken and PA methods for unidimen-

sional models. More detailed information about the performance of the methods for each condition by sample size (250 vs. 1000) is available online in Abdelhamid et al. (2024b), Supplementary Table S1. Overall, the AISP, GA and PApc achieved high accuracy ($\overline{pc}$ >.99), an absence of bias, and a low RMSD. In addition, both the AISP and GA presented a low rate of misclassification. On the other hand, EGA procedures obtained lower accuracy rates and higher misclassification rate and RMSD than AISP and GA. But, EGA with GLASSO and EGAtmfg with LE option performed better than EGA with GLASSO and EGAtmfg with the expand option.

Notably, EGA_expand with GLASSO showed better performance only on shorter scales (i.e., 5-item, 8-item, 10-item), and it performed poorly when the number of items was 16 or more, especially for the small sample size (250). EGAtmfg_expand only performed well with 5-item and 8-item factors, and it failed to identify unidimensional models when there were 10 items per factor or with 16 items or more, where $\overline{pc}$ = .00. The performance of both EGA with GLASSO and EGAtmfg with the LE option was better when the number of items was 20 or less, but performance degraded as the number of items increased, being poor when the number of items was 32 or more, especially for the small sample size (250).

In comparison with EGA with GLASSO and EGAtmfg, the performance of the AISP, GA and PApc was significantly less affected as the number of items per factor increased, where pc values were close to one and MR, MBE and RMSD being values close to zero.

Increasing the sample size from 250 to 1000 or 5000 improved the performance of the AISP, GA, EGA with GLASSO _expand, EGA_LE, and EGAtmfg_LE (see Abdelhamid et al., 2024b, Table S1 and Figure 2 and Figure 3. Regarding EGAtmfg_expand, its average performance did not improve by increasing the sample size from 250 to 1000 or 5000. Moreover, it is worth noting that the performance of PApaf decreased when the sample size was increased to 5000 in comparison to the performance achieved with sample sizes of 250 or 1000.

Figure 2

Proportion Correct, Misclassification Rate, Bias Error and Root Mean Squared Deviation of the Four Methods for Unidimensional Factors by the Independent Variables (Items per Factor and Sample Size)
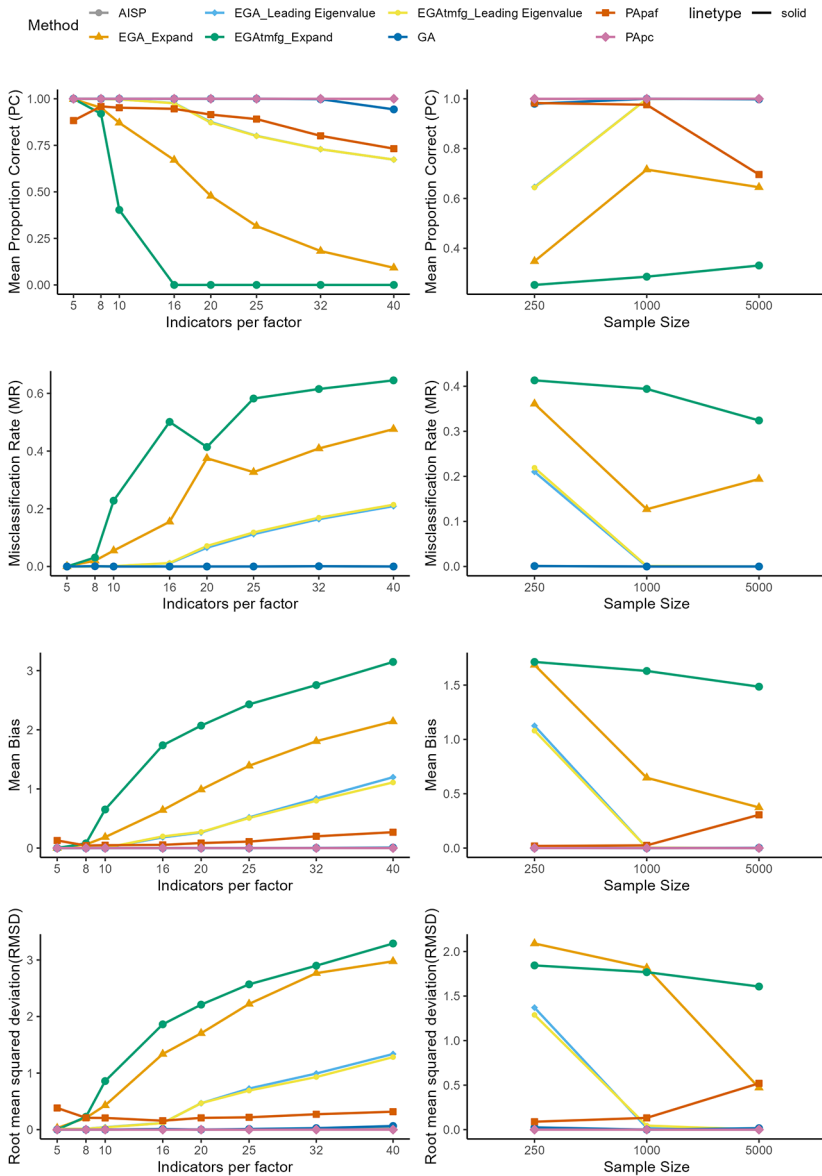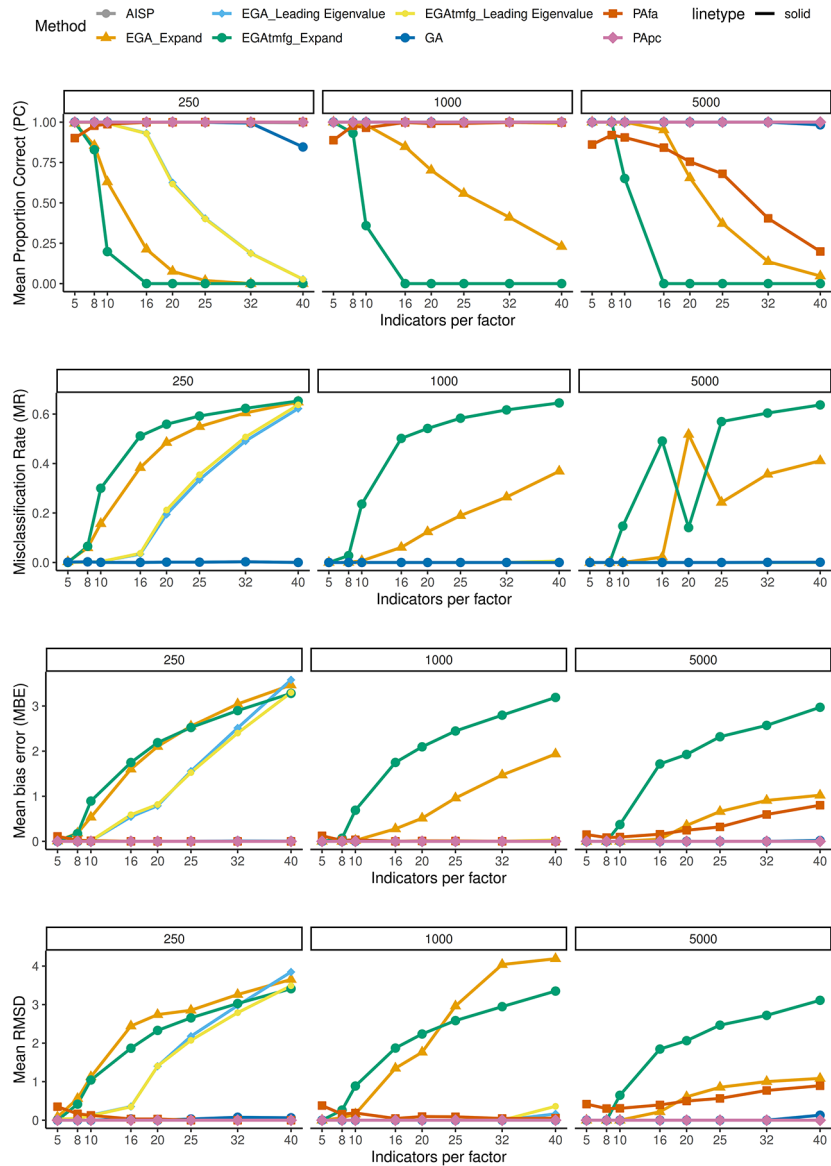
Figure 3

Proportion Correct, Bias Error, MR, and RMSD of the Four Methods for Unidimensional Factor Models Across the Levels of the Independent Variables

Multidimensional Factor Models

Figures 4, 5, 6, 7, and 8 summarize the performance of the four methods (EGA with GLASSO, EGAtmfg, AISP, GA, PApaf and PApc) in estimating the correct number of factors by sample size, number of factors, factor correlation, and number of items per latent factor. More detailed information about the performance of the four methods for each condition is available online in Abdelhamid et al. (2024b), Supplementary Tables S2–S4.

Overall, all methods were more accurate with orthogonal factor models ($r = 0$) than with correlated factor models. Regarding multidimensional factor models in general, the average accuracy for EGA with GLASSO,EGAtmfg, PApaf and PApc was higher than that for both the AISP and GA. In particular, EGA with GLASSO and EGAtmfg showed an improvement in their performance as the number of items per factor increased, and also as sample size increased. A similar pattern was found with the other three criteria: MR, MBE, and RMSD (see Figures 4, 5, 6, 7, and 8).

Notably, EGA with GLASSO and EGAtmfg exhibited the smallest overall $\overline{MR}$, followed by the AISP and GA. In addition, EGA with GLASSO showed the least bias, followed by PApaf and EGAtmfg. The two Mokken procedures, AISP and GA, yielded a larger MBE with under-factoring. EGA with GLASSO, EGAtmfg and PApaf were also the best methods when considering the RMSD criterion, values of which were higher with the GA, AISP and PApc.

Multiple Orthogonal Factors

With respect to multiple orthogonal factor models, EGA with GLASSO, EGAtmfg, PApaf, PApc and the AISP all achieved high accuracy, with an average of more than .99, whereas the average accuracy for the GA was only .789 (i.e., moderate accuracy). Concretely, EGA with GLASSO performed best with estimation of the correct number of factors with a smaller sample size ($n = 250$).

As shown in Figure 5, the performance of EGA with GLASSO, EGAtmfg, PApaf, PApc, and the AISP were improved when sample size increased from 250 to 1000 or 5000 across all conditions. Interestingly, the GA demonstrated worse results with an increasing number of factors, especially for conditions with more than four factors; its overall mean accuracy ($\overline{pc}$) for the two-factor, four-factor, and five-factor models was .998, .908, and .459, respectively.

Figure 4

Proportion Correct, Misclassification Rate, Bias Error and Root Mean Squared Deviation of the Four Methods for Multidimensional Factors by Four Independent Variables
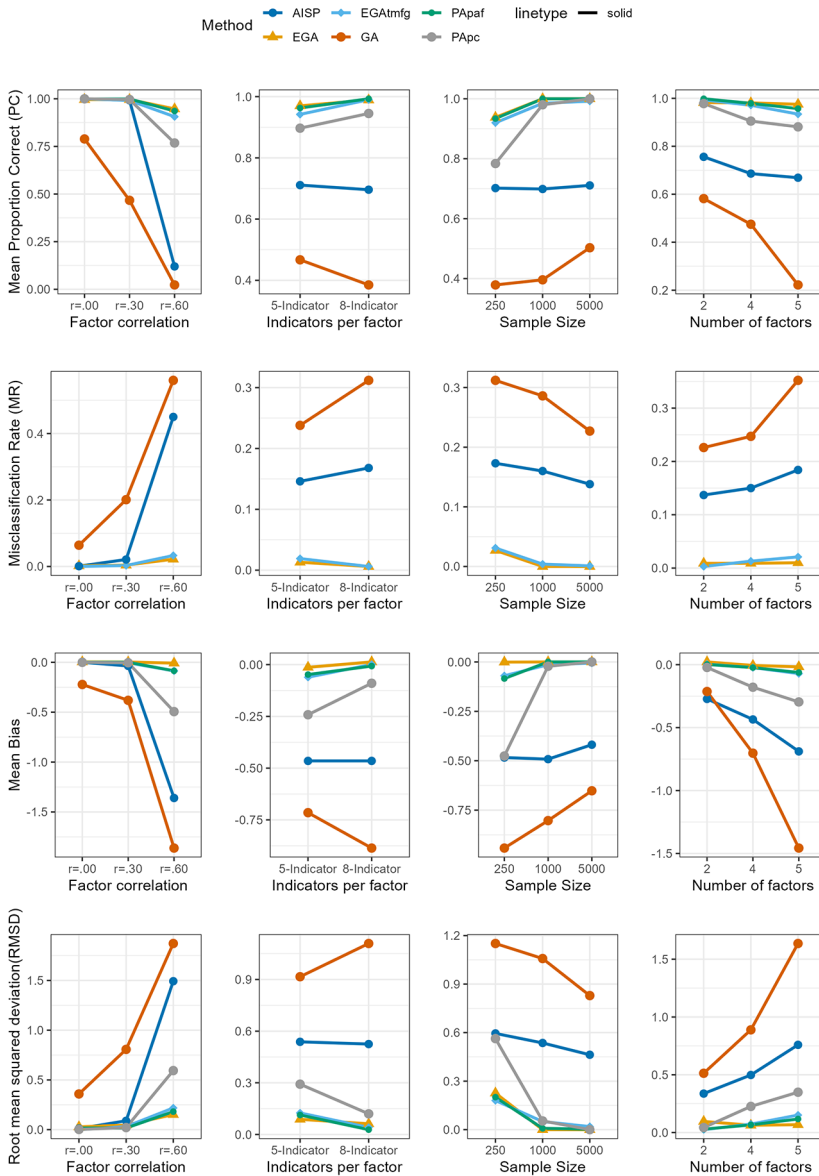
Figure 5

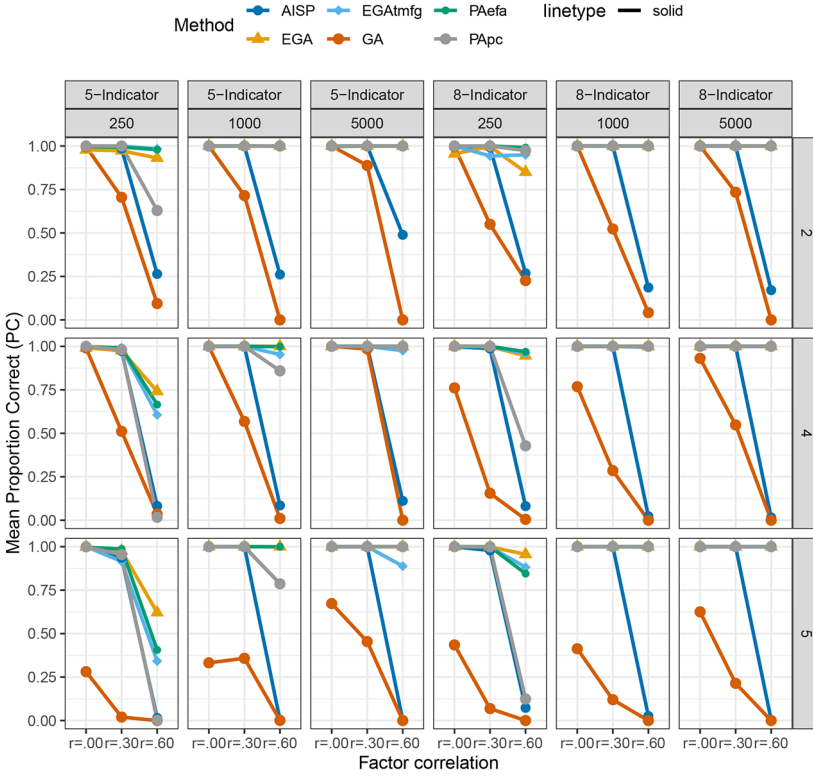*Mean Proportion Correct (PC) of the Four Methods for Multidimensional Factors Models*

Figure 6 displays the mean classification error rate of the four methods for the multidimensional factor model.

Figure 6 (left; $r = .00$) summarizes the MR results. These data show that EGA with GLASSO, EGAtmfg, PApaf, PApc, and the AISP performed well across all conditions regardless of the number of factors and sample size. However, this was not the case for the GA. It can be seen in Abdelhamid et al. (2024b), Supplementary Tables S2–S4, and Figure 6 that although the $\overline{MR}$ for the GA was very low when the number of factors was two ($\overline{MR} = .0012$), it increased, on average, to .038 and .151 for the four-factor and five-factor conditions, respectively. This indicates a high item misclassification rate in the case of five-factor models (around 15%).

Figure 6

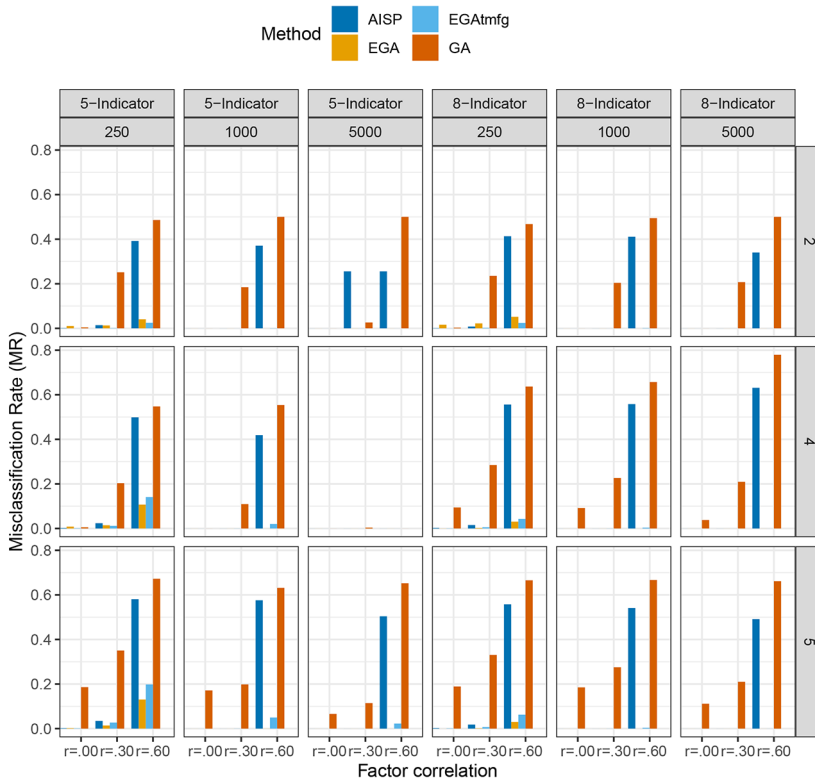*Mean Classification Error Rate of the Four Methods for Multidimensional Factors Models*

[Figure 7](#) depicts the bias results (MBE) for the six methods. EGAtmfg and PApc were unbiased across all conditions. Specifically, the bias associated with EGA with GLASSO, PApaf and the AISP decreased when sample size increased from 250 to 1000 or 5000. The GA yielded low bias across all conditions, except for the following: the five-factor model (under-factoring) and the four-factor model with 8 items per factor (under-factoring) (see also Tables S3 and S4, in [Abdelhamid et al., 2024b](#)).

With respect to the RMSD, and as shown in [Figure 4](#) and [Figure 8](#) (left; *r* = 0), EGAtmfg and PApc performed the best, followed by the PApaf, AISP, EGA with GLASSO, and the GA.

Figure 7

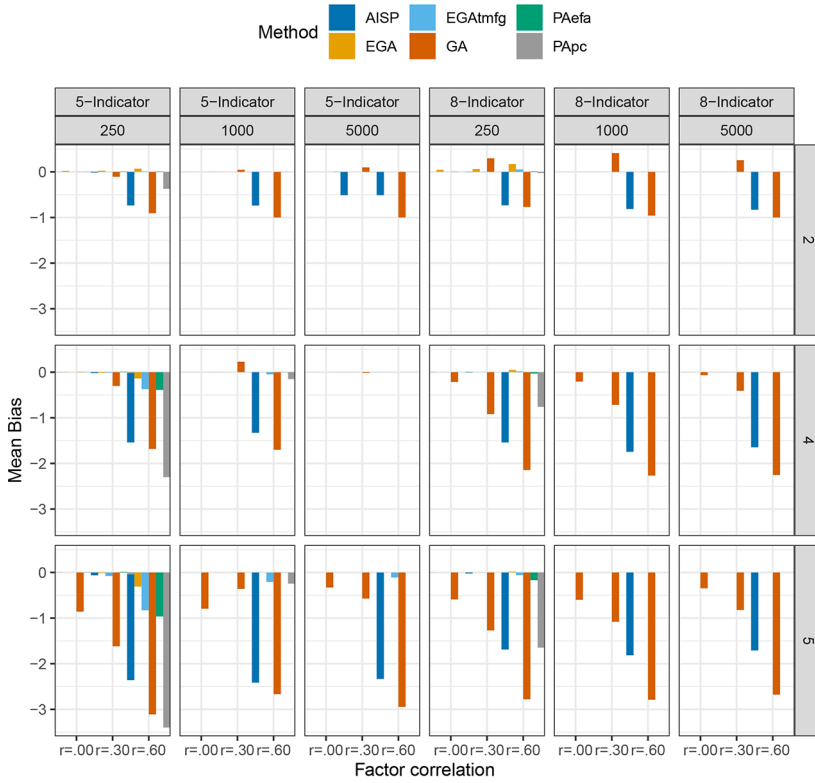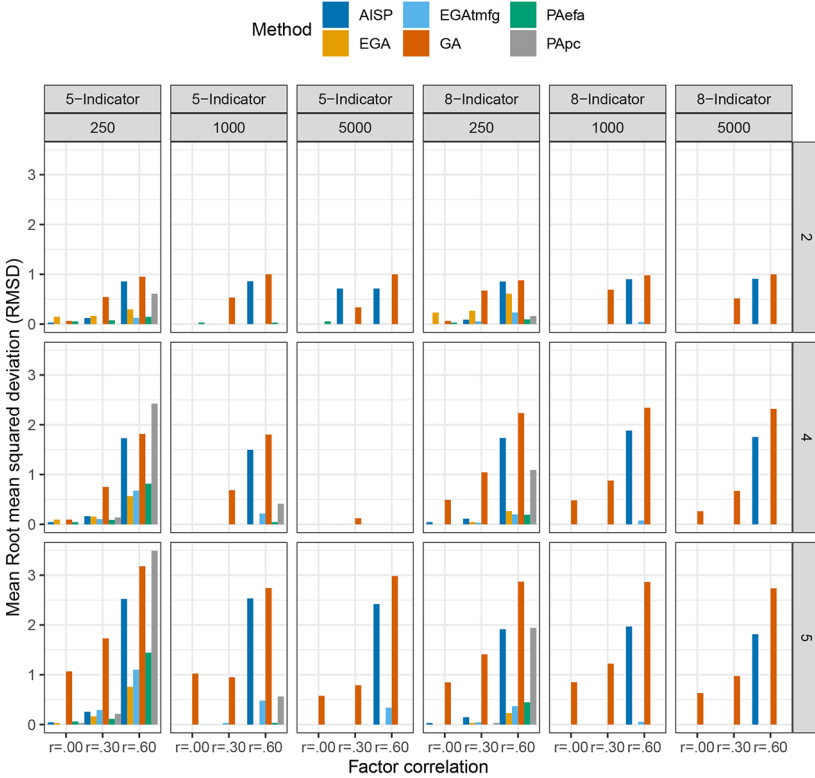*Mean Bias Error of the Four Methods for Multidimensional Factors Models*

PsychOpen GOLD

Figure 8

Mean Root Mean Squared Deviation of the Four Methods for Multidimensional Factors Models

Multiple Correlated Factors

Figures 4, 5, 6, 7, and 8 display the results for conditions with multiple correlated factors (r =.30 and .60). All methods were less accurate under these conditions than they were with multiple orthogonal factor models. Closer inspection of the data showed that the order of methods with multiple correlated factors was, from best to worst and on average: EGA with GLASSO, PApaf, EGAtmfg, PApc, AISP, and GA.

In conditions where the correlation between factors was r = .3 and with a small sample size (n = 250), EGA with GLASSO, EGAtmfg, PApaf, PApc and the AISP exhibited high accuracy (more than 90%) in retrieving the number of factors, regardless of the number of items per factor. In contrast, the GA showed low accuracy rates with a sample of 250. Further, *pc* rates decreased when the number of correlated factors increased. All methods showed under-factor identification in the conditions r = .3, n = 250, and 5 items per factor, with EGA with GLASSO being the least biased, followed by the PApaf, the

EGAtmfg, the PApc, the AISP, and the GA, respectively. When there were 8 items per factor, the correlation between factors was .3, and sample size was 250, PAefa, PApc and EGAtmfg showed the least bias, followed by EGA with GLASSO, the AISP, and the GA. Notably, when the sample size was large (n = 1000 or 5000), the performance of EGA with GLASSO, EGAtmfg, PApaf, PApc, and the AISP improved, whereas the performance of the GA remained poor.

Regarding the conditions with highly correlated factors(r = .6), EGA with GLASSO and PApaf displayed, on average, high accuracy, followed by EGAtmfg. The accuracy of the AISP and the GA was very low across all of these conditions. When the sample size was small (n = 250), accuracy was under .90 for EGA with GLASSO, EGAtmfg and PApaf, and very low for the AISP, the GA and PApc. Specifically, in conditions of four and five correlated factors with few items per factor (n = 5) and a small sample size (n = 250), all methods displayed low accuracy rates. For scales that included 8 items per factor and with a small sample size (n = 250), the accuracy of PApaf, PApc, EGA with GLASSO and EGAtmfg improved, whereas that of the AISP and the GA did not. Concretely, PApaf, EGA with GLASSO and EGAtmfg displayed high accuracy (above 85%) with a sample size of 1000 or 5000, regardless of the number of items per factor, except for the five-factor model with 5 items per factor in the case of EGAtmfg and PApc. By contrast, the performance of the AISP and the GA did not improve when sample size increased to 1000 or 5000.

Discussion

The assessment of true dimensionality for dichotomous data is a significant challenge in social and behavioral science research. The present study assessed the performance (in terms of *pc,* mean bias, RMSD, and MR) of four methods when applied to dichotomous items under several data conditions that are commonly found in psychological, sociological, and educational research. These methods were: EGA (using the GGM model) with GLASSO estimation, EGA with the TMFG algorithm, Mokken's original AISP, the GA, the PApaf, and the PAFpc. Using Monte Carlo simulation, an extensive assessment of these six methods was conducted by varying sample size, number of factors, number of items per factor, and factor correlations.

The clearest finding to emerge from the Monte Carlo simulation is that the performance of the six methods differed significantly according to data structure (i.e., unidimensional, orthogonal factors, and correlated factors). EGA with GLASSO estimation and PApaf yielded the greatest accuracy, the lowest bias, and the lowest rate of misclassification of items to factors in multidimensional structures (orthogonal factors and correlated factors), followed by EGAtmfg and PApc. Although the accuracy rates of EGA with GLASSO and EGAtmfg were high only for short unidimensional scales when unidimensionality was detected under the expand option. These findings are consistent

with the results reported by Golino et al. (2020) and Golino and Epskamp (2017). As for the performance of MSA methods, this was significantly affected by high correlation between factors. Mokken's original AISP performed well (high accuracy, low bias, and low rate of misclassification of items) when the scale structure was unidimensional or when correlations between factors were low in orthogonal factor and multiple correlated factor models. Unlike the other methods, the GA exhibited better performance only under limited conditions, namely unidimensional scales, bidimensional scales with orthogonal factors, and multidimensional short scales (5 items per factor) with four or fewer orthogonal factors. In accord with these findings is a consistent result with newer proposed methods that show slight advantages over similar methods used in this study, but the advantage of one method over another depends on the data conditions (e.g., Haslbeck & van Bork, 2024). That is, no method appears to be ideal for across all data conditions in simulation studies.

Regarding unidimensional scales, Golino et al. (2020) investigated the performance of the two EGA methods via expand with respect to a maximum of 12 dichotomous items per unidimensional scale, whereas the current simulation study examined a wide range of unidimensional scales, ranging from short (5 items) to long (40 items), via two methods (e.g., LE and expand). In general, the two methods of EGA with the LE option performed better than the expand option. We found that EGAtmfg with expand achieved high accuracy only for short scales consisting of five or eight items, and it yielded worse results for scales comprising 10 items or more regardless of sample size. This is in line with Golino et al. (2020), who found that EGAtmfg showed very low accuracy for 12-item conditions. As for EGA with GLASSO using the expand method, this method likewise only performed well with shorter scales, and lost accuracy when the number of items increased to 16 or more. These results would appear to be consistent with the findings of Golino et al. (2020), who reported that EGA with GLASSO using the expand showed high accuracy ($M = 92.54\%$) with a maximum of 12 variables per factor. Given the sample size of 250, both EGA with GLASSO and EGAtmfg showed higher pc and less MBE and misclassification rate with the LE option when the number of items was 16 or less. However, when the number of items was 25 or more, MBE and misclassification rate increased and pc decreased, such that these methods showed poor results. One interesting finding in the present study is that the performance of EGA was improved by increasing the sample size to 1000 or 5000, which endorses previous studies (e.g., Golino et al., 2020).

Regarding item selection in the context of MSA, Mokken's original AISP was better at retrieving correct factors than was the GA approach. This contrasts with the findings of Straat et al. (2013), who observed that the GA was a better method than the AISP. This discrepancy may be due to limitations in the design used by Straat and colleagues (i.e., the number of factors was limited to two, and only a large sample size of 1000 was considered). We found no evidence to support the results of Straat et al. (2013) with

regard to these conditions of two factors with 1000 subjects, especially for correlated factor scales. Importantly, we also found that the two Mokken methods did not retrieve the correct number of factors underlying scales with highly correlated factors. This is consistent with the results of Antino et al. (2018), who found that MSA formed only one scale regardless of sample size when the correlation between two factors was more than or equal to $r = .3$ and, in particular, from $r = .4$ onwards. These results partly corroborate the conclusions of Smits et al. (2012), who argued that the AISP was only useful for forming Mokken scales, where a single Mokken scale may be divided into multidimensional factors, and a single factor may consist of two Mokken scales. Furthermore, van Abswoude et al. (2004) noted that for test construction, it is often desirable to consider two dimensions that correlate .80 as a single dimension.

The current findings have important implications for partitioning dichotomous items into factors when using the Mokken methods (GA and AISP) and EGA procedures. The GA, based on our simulated conditions, in an inaccurate technique for partitioning a pool of items into correlated factors, and Mokken's original AISP should be used with caution, especially in the case of scales that contain highly correlated factors. Measures with high correlations between factors tend to form only one Mokken scale when using the AISP and the GA. Our results also suggest that EGA with GLASSO and EGAtmfg may be the gold-standard methods for partitioning multidimensional measures into factors. EGA with GLASSO, EGAtmfg and PApaf were able to retrieve the true structure of multidimensional scales, with high accuracy, low bias, a low rate of item misclassification on a factor, and low RMSD, regardless of the correlation between factors, sample size, number of items per factor, or number of factors. These results are in line with those found in previous work by Golino and Epskamp (2017) which showed that EGA is bit better than other procedures such as PA or CFA mainly at sample sizes of 500 or higher and with highly correlated number of Factors 4 and 6.

In general, the empirical results obtained in this research provide a basis for the use of EGA (with GLASSO), PApaf and EGAtmfg. Conversely, when analyzing unidimensional items or a multidimensional factor structure with low or zero correlations, AISP may be more suitable for recovering the latent structure. GA demonstrated good accuracy only under specific conditions, such as unidimensional scales, two-dimensional scales with orthogonal factors, and short multidimensional scales (5 items per factor) with four or fewer orthogonal factors. These findings are valuable for researchers and practitioners interested in assessing the dimensionality of a set of dichotomous items.

Despite all the technical and statistical aspects discussed in this paper, the researcher or practitioner cannot lose sight of the fact that the number of factors should not only be selected based on statistical considerations, but that the substantive theory, the nature of the construct being measured by the test, and the type of instrument being analyzed must be taken into account (Floyd & Widaman, 1995; Rosellini & Brown, 2021).

Limitations

The simulation conditions considered in this study are typical of those employed in research on identifying the dimensionality or number of latent factors, both in applied and simulated studies. However, no simulation will be exhaustive, and this research's simulation study is no exception. It is necessary to broaden the conditions to be examined.

In terms of limitations, the present study exclusively focused on dichotomous response variables. Further research is needed to determine the relative performance of these six methods with ordinal or continuous response variables. In addition, it should be noted that our study deliberately omitted the consideration of three factor when varying the number of true factors. This decision was made to avoid redundancy and maintain a manageable number of conditions. While a range of true factors between 1 and 4 is more conventional, our selection of 1, 2, 4, and 5 allowed us to effectively assess the impact of varying the number of true factors on the performance of different factor analysis methods. However, the exclusion of the Number 3 may be considered a limitation of our study. Future research could explore a broader range of true factors to provide a more comprehensive analysis.

Additionally, it is important to note that the model used in the present study to generate the data is essentially an alternative parameterization of a nonlinear multidimensional-UVA FA model. Consequently, the results obtained when fitting a nonlinear model to assess dimensionality and structure are expected to be superior to those obtained using any alternative model. Therefore, considering the data simulation model, it is expected that using MSA entails essentially a loss of information and effectiveness for assessing dimensionality and structure.

However, Mokken Scale Analysis and Exploratory Graph Analysis offer a different approach that can provide valuable insights on the dimensionality and structure of binary data under specific conditions.

Conclusion

The current simulation study investigated the performance of various methods for identifying the number of latent factors with dichotomous response items. The findings indicated that EGA (with GLASSO) and PApaf exhibited greater accuracy in general, followed by EGAtmfg. The AISP showed a high level of accuracy with unidimensional scales, orthogonal factor models, and multiple factor models with low correlations between factors. The GA only achieved good accuracy under limited conditions, namely unidimensional scales, bidimensional scales with orthogonal factors, and multidimensional short scales (5 items per factor) with four or fewer orthogonal factors. The results of this simulation study may serve as a guide for researchers when using EGA and MSA methods to partition dichotomous items into scales.

Supplementary Materials

For this article, the following Supplementary Materials are available:

- Additional information: Exploratory graph analysis and Mokken scale analysis (Appendix 1)—Abdelhamid et al., 2024a.
- Additional research results (Appendix 2)—Abdelhamid et al., 2024b.
- R syntax for EGA, Mokken, and PA (Appendix 3)—Abdelhamid et al., 2024c.

References

Abdelhamid, G. S. M., Gómez-Benito, J., Abdeltawwab, A. T. M., Abu Bakr, M. H. S., & Kazem, A. M. (2020). A demonstration of Mokken Scale analysis methods applied to cognitive test validation using the Egyptian WAIS-IV. *Journal of Psychoeducational Assessment, 38*(4), 493–506. https://doi.org/10.1177/0734282919862144

Abdelhamid, G. S. M., Hidalgo, M. D., French, B. F., & Gómez-Benito, J. (2024a). *Supplementary materials to* "Partitioning dichotomous items using Mokken scale analysis, exploratory graph analysis and parallel analysis: A Monte Carlo simulation" [Exploratory graph analysis and Mokken scale analysis]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.15425

Abdelhamid, G. S. M., Hidalgo, M. D., French, B. F., & Gómez-Benito, J. (2024b). *Supplementary materials to* "Partitioning dichotomous items using Mokken scale analysis, exploratory graph analysis and parallel analysis: A Monte Carlo simulation" [Tables of additional research results]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.15426/

Abdelhamid, G. S. M., Hidalgo, M. D., French, B. F., & Gómez-Benito, J. (2024c). *Supplementary materials to* "Partitioning dichotomous items using Mokken scale analysis, exploratory graph analysis and parallel analysis: A Monte Carlo simulation" [R syntax for EGA, Mokken, and PA]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.15427

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22*(3), 37–51. https://doi.org/10.1111/j.1745-3992.2003.tb00136.x

Ahmadi, K., Reidpath, D. D., Allotey, P., & Hassali, M. A. A. (2016). A latent trait approach to measuring HIV/AIDS related stigma in healthcare professionals: Application of Mokken scaling technique. *BMC Medical Education, 16*, Article 155. https://doi.org/10.1186/s12909-016-0676-3

Antino, M., Alvarado, J. M., Asún, R. A., & Bliese, P. (2018). Rethinking the exploration of dichotomous data: Mokken Scale analysis versus factorial analysis. *Sociological Methods & Research, 49*(4), 839–867. https://doi.org/10.1177/0049124118769090

Ayanwale, M. A., Amusa, J. O., Oladejo, A. I., & Ayedun, F. (2024). Multidimensional item response theory calibration of dichotomous response structure using R language for statistical computing. *Interchange, 55*(1), 137–157. https://doi.org/10.1007/s10780-024-09517-y

Ayanwale, M. A., Isaac-Oloniyo, F. O., & Abayomi, F. R. (2020). Dimensionality assessment of binary response test items: A non-parametric approach of Bayesian item response theory measurement. *International Journal of Evaluation and Research in Education, 9*(2), 385–393.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105*(3), 467–477. https://doi.org/10.1037/0033-2909.105.3.467

Chen, Y., & Zhang, S. (2021). Estimation methods for item factor analysis: An overview. In Y. Zhao & D. G. Chen (Eds.), *Modern statistical methods for health research: Emerging topics in statistics and biostatistics* (pp. 329–350). Springer. https://doi.org/10.1007/978-3-030-72437-5_15

Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36*(4), 523–562. https://doi.org/10.1207/S15327906MBR3604_03

Christensen, A. P., Garrido, L. E., Guerra-Peña, K., & Golino, H. (2024). Comparing community detection algorithms in psychometric networks: A Monte Carlo simulation. *Behavior Research Methods, 56*(3), 1485–1505. https://doi.org/10.3758/s13428-023-02106-4

Christensen, A. P., Kenett, Y. N., Aste, T., Silvia, P. J., & Kwapil, T. R. (2018). Network structure of the Wisconsin Schizotypy Scales–Short Forms: Examining psychometric network filtering approaches. *Behavior Research Methods, 50*(6), 2531–2550. https://doi.org/10.3758/s13428-018-1032-9

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed., pp. xii, 430). Lawrence Erlbaum Associates.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*, 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika, 82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x

Finch, W. H. (2020). Using fit statistic differences to determine the optimal number of factors to retain in an exploratory factor analysis. *Educational and Psychological Measurement, 80*(2), 217–241.

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*(4), 466–491.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286–299. https://doi.org/10.1037/1040-3590.7.3.286

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics, 9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods, 18*(4), 454–474. https://doi.org/10.1037/a0030005

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods, 21*(1), 93–111. https://doi.org/10.1037/met0000064

Golino, H., & Christensen, A. P. (2020). *EGAnet: Exploratory Graph Analysis – A framework for estimating the number of dimensions in multivariate data using network psychometrics.* https:// CRAN.R-project.org/package_EGAnet

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One, 12*(6), Article e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods, 25*(3), 292–320. https://doi.org/10.1037/met0000255

Haslbeck, J. M. B., & van Bork, R. (2024). Estimating the number of factors in exploratory factor analysis via out-of-sample prediction errors. *Psychological Methods, 29*(1), 48–64. https://doi.org/10.1037/met0000528

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393–416. https://doi.org/10.1177/0013164405282485

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

Immekus, J. C., Snyder, K. E., & Ralston, P. A. (2019). Multidimensional item response theory for factor structure assessment in educational psychology research. *Frontiers in Education, 4*, Article 45. https://doi.org/10.3389/feduc.2019.00045

Lauritzen, S. L. (1996). *Graphical models.* Oxford University Press.

Lee, H., & Cham, H. (2024). Comparing accuracy of parallel analysis and fit statistics for estimating the number of factors with ordered categorical data in exploratory factor analysis. *Educational*

and Psychological Measurement. Advance online publication.
https://doi.org/10.1177/00131644241240435

Li, Y., Jiao, H., & Lissitz, R. W. (2012). Applying multidimensional item response theory models in validating test dimensionality: An example of K–12 large-scale science assessment. *Journal of Applied Testing Technology, 13*(2), 103–110.

Lorenzo-Seva, U., & Ferrando, P. J. (2023). A simulation-based scaled test statistic for assessing model-data fit in least-squares unrestricted factor-analysis solutions. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 19*(2), 96–115. https://doi.org/10.5964/meth.9839

Lubbe, D. (2019). Parallel analysis with categorical variables: Impact of category probability proportions on dimensionality assessment accuracy. *Psychological Methods, 24*(3), 339–351. https://doi.org/10.1037/met0000171

Massara, G. P., Matteo, T. D., & Aste, T. (2016). Network filtering for big data: Triangulated maximally filtered graph. *Journal of Complex Networks, 5*(2), 161–178. https://doi.org/10.1093/comnet/cnw015

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research.* De Gruyter.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America, 103*(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

Nuño, L., Guilera, G., Barrios, M., Gómez-Benito, J., & Abdelhamid, G. S. M. (2022). Network analysis of the Brief ICF Core Set for schizophrenia. *Frontiers in Psychiatry, 13*, Article 852132. https://doi.org/10.3389/fpsyt.2022.852132

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications, 10*(2), 191–218. https://doi.org/10.7155/jgaa.00124

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reckase, M. D. (2009). *Multidimensional item response theory.* Springer Science & Business Media.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27–48.

Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research* [R Package Version 2.4.6]. Northwestern University. https://CRAN.R-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Rosellini, A. J., & Brown, T. A. (2021). Developing and validating clinical questionnaires. *Annual Review of Clinical Psychology, 17*, 55–81. https://doi.org/10.1146/annurev-clinpsy-081219-115343

PsychOpen GOLD

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45*(1), 73–103. https://doi.org/10.1080/00273170903504810

Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical & Statistical Psychology, 66*(2), 201–223. https://doi.org/10.1111/j.2044-8317.2012.02049.x

Sijtsma, K., Emons, W. H. M., Bouwmeester, S., Nyklíček, I., & Roorda, L. D. (2008). Nonparametric IRT analysis of quality-of-life scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref). *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 17*(2), 275–290. https://doi.org/10.1007/s11136-007-9281-6

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory* (1st ed.). SAGE.

Smits, I. A. M., Timmerman, M. E., & Meijer, R. R. (2012). Exploratory Mokken scale analysis as a dimensionality assessment tool: Why scalability does not imply unidimensionality. *Applied Psychological Measurement, 36*(6), 516–539. https://doi.org/10.1177/0146621612451050

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2013). Comparing optimization algorithms for item selection in mokken scale analysis. *Journal of Classification, 30*(1), 75–99. https://doi.org/10.1007/s00357-013-9122-y

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2014). Minimum sample size requirements for Mokken scale analysis. *Educational and Psychological Measurement, 74*(5), 809–822. https://doi.org/10.1177/0013164414529793

Svetina, D., & Levy, R. (2012). An overview of software for conducting dimensionality assessment in multidimensional models. *Applied Psychological Measurement, 36*(8), 659–669. https://doi.org/10.1177/0146621612454593

Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological Assessment, 31*(12), 1442–1455. https://doi.org/10.1037/pas0000597

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267–288.

van Abswoude, A. A. H., van der Ark, L. A., & Sijtsma, K. (2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement, 28*(1), 3–24. https://doi.org/10.1177/0146621603259277

van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software, 48*(5), 1–27.

van der Ark, L. A., Croon, M. A., & Sijtsma, K. (2008). Mokken scale analysis for dichotomous items using marginal models. *Psychometrika, 73*(2), 183–208. https://doi.org/10.1007/s11336-007-9034-z

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* (pp. 41–71). Springer US. https://doi.org/10.1007/978-1-4615-4397-8_3

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28*(3), 263–311. https://doi.org/10.1207/s15327906mbr2803_1

Wind, S., & Wang, Y. (2023). Using Mokken scaling techniques to explore carelessness in survey research. *Behavior Research Methods, 55*(7), 3370–3415. https://doi.org/10.3758/s13428-022-01960-y

Wismeijer, A. A. J., Sijtsma, K., van Assen, M. A. L. M., & Vingerhoets, A. J. J. M. (2008). A comparative study of the dimensionality of the self-concealment scale using principal components analysis and Mokken scale analysis. *Journal of Personality Assessment, 90*(4), 323–334. https://doi.org/10.1080/00223890802107875