Dimensionality assessment in generalized bi-factor structures: a network psychometrics approach

Marcos Jimenez[1] Francisco J. Abad[1]

Eduardo Garcia-Garzon[2] Hudson Golino[3]

Alexander P. Christensen[4] Luis Eduardo Garrido[5]

[1]Universidad Autónoma de Madrid (Spain)
[2]Universidad Camilo José Cela (Spain)
[3]University of Virginia (United States)
[4]Vanderbilt University (United States)
[5]Pontificia Universidad Católica Madre y Maestra (Dominican Republic)

ORCID

Marcos Jimenez https://orcid.org/0000-0003-4029-6144
Francisco J. Abad http//orcid.org/0000-0001-6728-2709
Eduardo Garcia-Garzon https://orcid.org/0000-0001-5258-232X
Hudson Golino https://orcid.org/0000-0002-1601-1447
Alexander P. Christensen https://orcid.org/0000-0002-9798-7037
Luis Eduardo Garrido https://orcid.org/0000-0001-8932-6063

Corresponding author: Eduardo Garcia-Garzon

Universidad Camilo José Cela

Villafranca del Castillo

28692 Madrid (Spain)

Email: egarcia@ucjc.edu

Author note

Funding: This research was supported by Grant PSI2017-85022-P (Ministerio de Ciencia, Innovación y Universidades, Spain) and the UAM IIC Chair Psychometric Models and Applications. Luis Eduardo Garrido is supported by Grant 2018-2019-1D2-085 from the Fondo Nacional de Innovación Desarrollo Científico y Tecnológico (FONDOCYT) of the Dominican Republic.

Conflicts of interest: None.

Material and data availability: The `bifactor` package is available at https://github.com/Marcosjnez/bifactor and all the files necessary to reproduce the simulation data, analyses, and figures can be found at https://osf.io/u7qwj/.

Abstract

The accuracy of factor retention methods for structures with one or more general factors, typically encountered in fields like intelligence, personality, and psychopathology, has often been overlooked in dimensionality research. To address this issue, we compared the performance of several factor retention methods in this context, including a new, highly accurate network psychometrics approach. For estimating the number of group factors, these methods were the Kaiser criterion, empirical Kaiser criterion, parallel analysis with principal components (PA_{PCA}) or principal axis, and exploratory graph analysis with Louvain clustering (EGA_{LV}). We then estimated the number of general factors using the factor scores of the first-order solution suggested by the best two methods, yielding a "second-order" version of PA_{PCA} ($PA_{PCA\text{-}FS}$) and EGA_{LV} ($EGA_{LV\text{-}FS}$). Additionally, we examined the direct multi-level solution provided by EGA_{LV}. All the methods were evaluated in an extensive simulation manipulating nine variables of interest, including population error. The results indicated that EGA_{LV} and PA_{PCA} displayed the best overall performance in retrieving the true number of group factors, the former being more sensitive to high cross-loadings, and the latter to weak group factors and small samples. Regarding the estimation of the number of general factors, both $PA_{PCA\text{-}FS}$ and $EGA_{LV\text{-}FS}$ showed a close to perfect accuracy across all the conditions, while EGA_{LV} was inaccurate. The methods based on EGA were robust to the conditions most likely to be encountered in practice. Therefore, we highlight the particular usefulness of EGA_{LV} (group factors) and $EGA_{LV\text{-}FS}$ (general factors) for assessing bi-factor and generalized bi-factor structures.

Keywords: *Dimensionality Assessment, Exploratory Factor Analysis, Exploratory Graph Analysis, Generalized Bi-factor, Parallel Analysis*

1 Introduction

Dimensionality assessment plays a central role in psychometrics, as it constitutes one of the cornerstone decisions during test validation. Unfortunately, simulation studies that focus on bi-factor structures are scarce in dimensionality research, and there are few recommendations on how to proceed when assessing the dimensionality of these structures. This comes as a surprise given the current popularity of bi-factor models in fields like intelligence (Beaujean, 2015), personality (Abad et al., 2018), and psychopathology (Bornovalova et al., 2020). The main feature of bi-factor models is that items are allowed to simultaneously load on a collection of group factors (e.g., generosity and tolerance) and one general factor (e.g., agreeableness), with the group factors representing narrower traits that explain the residual common variance that is left after extracting the general factor (Reise, 2012). Although the development of exploratory bi-factor techniques is still an active line of research, with proposals involving analytic rotation criteria (Jennrich & Bentler, 2011) and target-based procedures (Abad et al., 2017; Garcia-Garzon et al., 2019), they have been generalized to cover more than one general factor. Some examples are the two-tier hierarchical model of Tian and Liu (2021) and the generalized bi-factor model of Jimenez et al. (2022; Figure 1). These generalizations have the advantage of estimating several bi-factor structures in a single model, uncovering relationships that would remain hidden if we performed independent bi-factor analyses for each domain of the factor structure (e.g., correlations and cross-loadings among the general factors would not be estimated in these cases).

Generalized bi-factor models reflect the consensus that many psychological phenomena are hierarchically organized, with lower-level traits being subsumed in broader, more general factors. In fact, there have already been some efforts to explore and test these hierarchical organizations, such as the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017; Ringwald et al., 2021), which is a dimensional alternative to the Diagnostic and Statistical Manual of Mental Disorders (DSM) that conceptualizes psychopathology across different strata, namely symptoms, syndromes, sub-factors, and spectra.

Despite previous advancements in (generalized) exploratory bi-factor analysis, their application still requires a decision regarding the number of group and general factors to extract. Unfortunately, simulation studies including general factors are scarce and usually focus on structures with second-order general factors instead of on generalized bi-factor structures. Bi-factor models are only equivalent to second-order structures when proportionality constraints between the group and general factors are satisfied (Mansolf & Reise, 2016), so simulations covering the specific bi-factor case are required. In this context, some researchers have already investigated the behavior of parallel analysis methods in bi-factor structures to recover the number of group factors (Crawford et al., 2010; Green et al., 2015, 2016; Levy et al., 2021). However, the extent to which other factor retention methods work for this purpose is unknown. Furthermore, the quality of the recovery of the number of general factors is largely untested.

As such, the aim of this study was two-fold: firstly, investigating the capability of some popular factor retention methods to uncover the number of group factors in bi-factor and generalized bi-factor structures. These methods were the Kaiser criterion (K1), the empirical Kaiser criterion (EKC), parallel analysis with principal component analysis (PA$_{\text{PCA}}$), parallel analysis with principal axis factoring (PA$_{\text{PAF}}$), and exploratory graph analysis with the Louvain clustering algorithm (EGA$_{\text{LV}}$). The second goal of the study involved testing the performance of the methods to detect the number of general factors. While EGA$_{\text{LV}}$ automatically estimates a multi-level configuration of the data, another possibility consists in applying any of the aforementioned factor retention methods to the estimated first-order factor scores derived from an exploratory factor analysis (EFA). This second-order approach was performed using EGA$_{\text{LV}}$ and PA$_{\text{PCA}}$, giving rise to EGA$_{\text{LV-FS}}$ and PA$_{\text{PCA-FS}}$, respectively.

2 Dimensionality assessment methods

2.1 The Kaiser Criterion

The Kaiser criterion (K1; Kaiser, 1960), also known as the Kaiser eigenvalue-greater-than-one criterion, is one of the first and most popular factor retention methods. According to K1, the first k greater-than-one eigenvalues of a correlation matrix are indicative of k factors. This criterion was devised under the rationale that substantive factors should explain at least more variance than the average variance of the variables, which is one for correlation matrices, and to prevent the estimated factors from having negative reliability (Cliff, 1988). However, K1 gives an asymptotic lower bound for the number of true dimensions (Guttman, 1954). At the sample level, its low accuracy has been replicated by a large body of simulation research (Auerswald & Moshagen, 2019; Ruscio & Roche, 2011; Yeomans & Golder, 1982; Zwick & Velicer, 1986). The poor performance of K1 can be attributed to the bias of the sample eigenvalues. The first sample eigenvalue (λ_1) is the maximum value obtained from the optimization problem $\underset{\mathbf{x}\in\mathbb{S}}{\operatorname{argmax}}\ \mathbf{x}^\top\mathbf{S}\mathbf{x}$, where $\mathbf{S}$ is the sample correlation matrix and $\mathbf{x}$ is the first eigenvector of $\mathbf{S}$ (e.g., a vector estimated from the set of all unit vectors, $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^p : \langle \mathbf{x}, \mathbf{x} \rangle = 1\}$). Subsequent eigenvalues are estimated similarly, but constraining the new estimated vectors to remain orthogonal to all the previous ones. This serial dependency results in the first sample eigenvalues being upwardly biased, as they have more variance to capitalize on by chance with fewer constraints. Thus, the bias of the sample eigenvalues is inversely related to the sample size and positively related to the number of variables, as there is more noise in small samples with a large number of variables, leading K1 to overestimate the true number of factors.

However, learning this important shortcoming has not prevented the widespread use of K1. Goretzko et al. (2021) reviewed the exploratory factor analysis literature published between 2007 and 2017 in two psychological journals with a special focus on test development and found that K1 was the most common method either when several factor retention methods were simultaneously used (55.6%) and when a single method was used (10.5%). To our knowledge, the performance of K1 has not been investigated in the presence of general factors. Therefore, we decided to include it in our simulations to further investigate its problematic capitalization on chance in bi-factor and generalized bi-factor structures.

2.2 The Empirical Kaiser Criterion

Braeken and Assen (2017) proposed the Empirical Kaiser Criterion (EKC), a modification of K1 that considers the serial dependency between the sample eigenvalues. EKC compares the sample eigenvalues to reference eigenvalues (λ^{EKC}) that are sequentially computed under a null model with no latent factors. Asymptotically, if the variables are normally distributed, the eigenvalues of the sample correlation matrix follow the Marčenko-Pastur distribution (Marčenko & Pastur, 1967). Hence, Braeken and Assen (2017) chose the first reference eigenvalue under the null model (λ_1^{EKC}) to equal the expected value of the first sample eigenvalue from the Marčenko-Pastur distribution,

$$\lambda_1^{EKC} = (1 + \sqrt{\gamma})^2, \quad \text{with}\, \gamma = J/n, \tag{1}$$

where n is the sample size and J is the number of variables. The subsequent reference eigenvalues λ_j^{EKC}, $j = \{2, 3, \ldots, J\}$, are then computed multiplying (1) by the average variance that is left after taking out the first $j - 1$ factors, $(J - \sum_{j=0}^{j-1} \lambda_j)/(J - j + 1)$, where $\lambda_0 = 0$. The resulting reference eigenvalues can then be interpreted as an estimate of the population value of λ_j if the null model of conditional independence was true after accounting for $j - 1$ factors.

Overall, the formula for computing the reference eigenvalues is

$$\lambda_j^{EKC} = \max\left(\frac{J - \sum_{j=0}^{j-1} \lambda_j}{J - j + 1}(1 + \sqrt{\gamma})^2, 1\right). \tag{2}$$

Notice that the minimum reference eigenvalue is set to one to guarantee that, at the population level, K1 and EKC match in the number of factors to retain, representing a lower bound for the true number of factors (Guttman, 1954).

EKC has been suggested to be more robust than parallel analysis in conditions involving few variables per factor and high factor correlations (Auerswald & Moshagen, 2019; Braeken & Assen, 2017) and in the presence of cross-loadings in multivariate normal data (Li et al., 2020). However, its performance has not been tested in bi-factor and generalized bi-factor structures. We thus included this recent proposal in our simulation study.

2.3 Parallel Analysis

Parallel analysis (PA; Horn, 1965) has been considered the gold-standard method for dimensionality assessment for many decades, with many simulation studies recommending its use for either continuous (Fabrigar et al., 1999; Lim & Jahng, 2019; Zwick & Velicer, 1986) and ordinal data (Garrido et al., 2016, 2013; Timmerman & Lorenzo-Seva, 2011). PA would emulate the sampling process of the original correlation matrix if no latent factors were present, controlling the impact that the sample size and the number of variables bear in the magnitude of the eigenvalues. More specifically, PA compares the sample eigenvalues to reference eigenvalues obtained by simulating data from a null model, with the first k sample eigenvalues greater than their corresponding reference eigenvalues being indicative of k meaningful factors.

The reference eigenvalues can be computed in many ways. In the original formulation, Horn (1965) performed principal component analysis in a large number of $n \times J$ matrices of uncorrelated normally distributed random variables, using the average of the empirical distribution of the eigenvalues as the reference eigenvalues. Later proposals involved the use of the 95th percentile of the empirical distribution instead of the mean (Buja & Eyuboglu, 1992; Glorfeld, 1995), the resampling of the observed data matrix for generating new random data (PA$_{\text{PCA}}$; Buja & Eyuboglu, 1992), the replacement of principal components either by principal axis factoring (PA$_{\text{PAF}}$; Humphreys & Ilgen, 1969) or minimum rank factor analysis

(Timmerman & Lorenzo-Seva, 2011), and the assessment of each j factor in a sequential manner, taking the $j-1$ factor model as the referent null model for generating random data (Green et al., 2012).

Several simulation studies comparing different versions of PA have found that even though no single method outperformed others in all conditions, PA_{PCA} presented the highest overall accuracy (Lim & Jahng, 2019; Xia, 2021). However, other authors support employing PA_{PAF} instead, arguing that it outperforms PA_{PCA} under conditions with multiple correlated factors (Crawford et al., 2010; Keith et al., 2016). In the particular case of structures including general factors (in both second-order and bi-factor structures), Crawford et al. (2010) found that PA_{PCA} tended to recover the number of general factors while PA_{PAF} accurately recovered the number of group factors. However, Lim and Jahng (2019) and Xia (2021) noted that this superiority vanishes when the realistic condition of population error is included. This current controversy prompted the examination of both methods in our simulations.

Other authors have also contended that Revised parallel analysis (RPA) is superior to PA_{PCA} in structures involving both correlated first-order factors and bi-factor structures (Green et al., 2016, 2015). RPA was developed when Green et al. (2012) acknowledged that the random data generation of parallel analysis does not take into account the sequential nature of eigenvalues and argued that a fairer comparison between the kth sample and kth reference eigenvalue is attained when the latter is computed using random data generated from an estimated $k-1$ factor model. However, their simulations did not consider the role of population error. Lim and Jahng (2019) and Xia (2021) also tested RPA in their simulations and concluded that PA_{PCA} was superior to RPA when a sufficient misfit was added to the population correlation matrix. Cosemans et al. (2021) also found that RPA was very sensitive to sample size, with larger samples severely undermining its accuracy. In the context of generalized bi-factor structures, another disadvantage is that RPA would be time-consuming for these large structures. Therefore, we did not include this version of parallel analysis in our simulation.

Finally, concerning the cut-off value needed to derive the reference eigenvalues, Xia (2021) showed that the performance of PA_{PCA} using the $95th$ percentile was more robust to model misspecification than the mean value. In contrast, the mean of the empirical eigenvalues was more robust to multiple correlated factors. These results are explained by stringent cut-offs ignoring minor factors and larger cut-offs avoiding the collapse of correlated factors. In the end, as our simulations included model error and at the same time the group factors were correlated due to the presence of the general factors, we decided to conduct PA_{PCA} and PA_{PAF} with both the mean and the $95th$ percentile.

2.4 Exploratory Graph Analysis

Network psychometrics is an alternative method to factor analysis to model and interpret psychological data. In a network model, a random variable is a node connected to other nodes by edges typically representing their relationship after conditioning on all the other variables. In the same way that factor models are commonly displayed with diagrams, networks models are visualized with a graph containing all the nodes and edges connecting them, with nodes belonging to the same cluster being placed closer, and edge's thickness representing the strength of the associations between the nodes (Figure 2).

For multivariate normal data, the most straightforward way to model such pairwise relationships among the variables is using their partial correlations. This is the simplest way of estimating a Gaussian Graphical Model (GGM; Epskamp et al., 2018). However, Epskamp and Fried (2017) warned that when two variables are conditionally independent, the partial correlation matrix usually reflects spurious relationships due to sampling variation, leading to large standard errors and unstable parameter estimates. As a solution, regularization techniques such as the graphical least absolute shrinkage and selection operator (GLASSO; Friedman et al., 2008) are used to estimate sparse partial correlations. GLASSO regularization contains a tuning parameter controlling the sparsity of the network that is selected by minimizing a complexity function such as the Extended Bayesian Information Criterion

(EBIC; Chen & Chen, 2008). With this approach, small partial correlations are shrunk towards zero, yielding a more parsimonious and interpretative network with more unconnected nodes reflecting conditional independence. Latent factors underlying the data can then be related to clusters of nodes, with edges within a cluster being stronger than between clusters (Golino & Epskamp, 2017). Such reciprocity between clusters of nodes and latent variables is not only justified by the fact that network models are statistically consistent with factor models under certain conditions (Bork et al., 2021) but also supported by empirical research and simulation studies (Golino & Demetriou, 2017; Golino, Shi, et al., 2020).

Network psychometrics settles the grounds for Exploratory Graph Analysis (EGA; Golino & Epskamp, 2017) as a factor retention method. Firstly, EGA estimates the partial correlations between the variables by fitting a GGM with the GLASSO regularization and then applies a community detection algorithm for weighted networks to classify items into clusters. Usually, the clustering is achieved by maximizing *modularity*, an index measuring the extent to which nodes within a cluster are more connected than between clusters. Christensen et al. (2020) performed a simulation comparing eight clustering algorithms and found that the Louvain (Blondel et al., 2008) and Walktrap (Pons & Latapy, 2006) algorithms (both based on modularity) attained the best overall results in identifying the true number of dimensions. For our simulation study, we decided to implement EGA with the Louvain algorithm (EGA_{LV}) because it performs at least as well as the Walktrap algorithm and potentially provides multi-level clusters. Christensen et al. (2020) suggested that *"preference should be given to the Louvain algorithm because it also provides a hierarchical or "multi-level" structuring of factors. Such hierarchical structuring is important for determining different levels of taxonomies that often exist in psychological assessment instruments"* (p. 18). Thus, the Louvain algorithm can directly estimate the number of general factors, although it has not been used yet for this purpose in psychological data. Notwithstanding, a drawback attached to modularity optimization algorithms is that the modularity index is zero for unidimensional structures, so they tend to estimate more than one factor in these cases. As a solution, Christensen et al. (2020) incorporated to EGA an initial check using the Leading eigenvalue community detection algorithm (LE; Newman, 2006) on the raw correlation matrix. LE also maximizes modularity but was considered to provide an adequate balance between correctly recovering one and more than one factors. As such, if LE delivered one factor, the data was judged to be unidimensional. Contrary, when it estimated more than one factor, EGA_{LV} was applied instead.

To date, no exploratory graph analysis method has been tested in bi-factor and generalized bi-factor structures despite the appealing multi-level feature of the Louvain algorithm. Thus, we included it in our simulation.

3 Assessing the number of general factors

If the number of group factors and their configural structure were known, we could roughly estimate the number of general factors by summing or averaging the items corresponding to each scale and then employing any previous factor retention method over the resulting scores. However, this strategy is unrealistic because the first-order dimensionality and the factor pattern are often unknown or unclear. One alternative is Goldberg's Bass-Ackwards method (Goldberg, 2006), a sequential top-down approach that starts by estimating a unidimensional exploratory factor model and continues extracting and rotating factors until no variable primarily loads on a factor. Then, the factor scores for each factor solution are estimated, and their correlations are used to build a hierarchical representation of all the factor solutions, with the first-factor solution depicted at the top, followed by the two rotated factors solution, and so on. Then, high correlations between an upper and a lower-order factor indicate the perpetuation of the factor down the hierarchy. In contrast, medium correlations between a certain upper and lower-order factor indicate that the former was split to yield the latter, narrower factor.

Notwithstanding, the Bass-Ackwards method rests on a top-down approach, assessing first the higher-order factors in the hierarchy. Condon et al. (2020) warned that top-down approaches are at risk of missing important features of the factor structure. For instance, they are unable to identify the presence of gaps in content concerning the higher-order domains and are also susceptible to the jingle-jangle fallacy (e.g., we are at risk of labeling with different names the same trait down the hierarchy (jingle) and using the same label for different traits (jangle)). In contrast, they argue for a bottom-up approach that starts by assessing all the traits or nuances that exhaust a domain, taking into account item complexity and facilitating item revision and content expansion.

An example of a bottom-up approach is the one proposed by Golino, Jotheeswaran, et al. (2020). First, estimated the number of group factors using EGA. Secondly, they estimated a loading matrix for the group factors from the fitted network and obliquely rotated the structure employing geomin. Finally, they used the resulting first-order latent factor correlation matrix to perform a second-order EGA, yielding an estimation of the number of general factors. However, this procedure was developed to investigate the relationship between several cognitive and health-related variables in the context of aging research, and no exhaustive simulation was performed to test its accuracy under different scenarios of interest.

In this study, we also followed a bottom-up method based on the correlation between the factor scores of the group factors, as they are expected to reflect the latent dependencies between the general factors. We would like to remark that we are not the first in suggesting nor using factor scores from lower-order factors to determine the number of general factors (see Friborg et al., 2009 and Milfont & Duckitt, 2004). However, previous proposals were not fully explicit or included steps that did not align with what we think would be best practice (e.g., using composites of items for estimating the factor scores, performing orthogonal rotation, or using K1 to assess the number of general factors). The solution that we propose is straightforward and can be obtained through the following steps: (a) estimate the number of group factors with some factor retention method; (b) perform an oblique exploratory factor analysis of the observed correlation matrix extracting the number of group factors suggested in the previous step; (c) estimate the factor scores with some method that contemplates correlated factors (e.g., Thurstone's regression method); and (d) estimate the number of general factors on the factor scores using the same factor retention method employed in the first step. In our simulations, the methods that followed this second-order procedure, first estimating the number of group factors on the item scores and then estimating the number of general factors on the factor scores, were performed using PA_{PCA} and EGA_{LV} , yielding $PA_{PCA\text{-}FS}$ and $EGA_{LV\text{-}FS}$.

4 Methods

4.1 Simulation design

Following a similar design to those found in Abad et al. (2017) and Jimenez, Abad, Garcia-Garzon, & Garrido (2022), nine variables were manipulated to create realistic bi-factor and generalized bi-factor structures: (a) number of general factors (N.GF: 1, 2, 3); (b) correlation between the general factors (COR.GF: 0, .30); (c) sample size (N: 500, 1000, 2000, 5000); (d) number of group factors per general factor (NUM.GRF: 4, 5, 6); (e) number of variables per group factor (VAR.GRF: 4, 6, 8, 10); (f) factor loadings on the general factors (LOAD.GF: low, medium); (g) factor loadings on the group factors (LOAD.GRF: low, medium); (h) model error or misfit (MF: zero, close); and (i) cross-loadings among the group factors (CROSS.GRF: 0, .15, .30). These variables were crossed to yield a final number of 5760 conditions, after removing the incompatible conditions in which the number of general factors was set to one but the correlation between the general factors was not zero.

Factor loadings ranged from .30 to .50 for the low condition and from .40 to .60 for the medium condition. The loadings on the general factors were sampled from a uniform distribution, whereas the loadings on the group factors varied by equal increments across their variables (e.g., for the low condition with four items per group factor, the population factor loadings were .30, .37, .43, and .50). To create conditions with cross-loadings, the item with the greatest loading on each group factor had a cross-loading of .15 or .30 in another group factor. Then, we maintained the communality constant by subtracting a small value from the remaining non-zero item loadings to make the conditions with and without cross-loadings comparable (see Abad et al., 2017). To illustrate how the data were simulated under these conditions, Table 1 shows a randomly generated loading pattern matrix corresponding to a bi-factor model before and after introducing the cross-loadings. Generalized bi-factor structures with more than one general factor were created by simply joining these single bi-factor structures row-wise.

4.2 Population misfit

In real situations, the population correlation matrix between the variables does not resemble the correlation matrix reproduced by the true model parameters due the to population model misfit (MacCallum, 2003). In other words, all models are misspecified because of many unmodeled minor factors explaining some common item variance. According to this perspective, the true number of factors underlying a population correlation matrix corresponds to the number of major factors, and the resulting population misfit is interpreted as a trivial nonsubstantive common variance. In our simulations, population misfit was created following the method proposed by Cudeck and Browne (1992). This method generates small random values that are added to the population implied correlation matrix such that fitting a confirmatory factor model with unweighted least squares (ULS) reproduces the intended amount of misfit while preserving a global minimum at the original model parameters, as long as the error is not excessive.

We selected the population standardized root mean square residual (SRMR) as the indicator of the amount of global misfit, following Shi et al. (2018) and Ximénez et al. (2022). The former investigated the behavior of the population SRMR under different types and degrees of model misspecification to suggest a corrected cut-off for the population SRMR that corresponds to a close-fitting model. They established that a close-fitting model at the population level exists when (1) the largest absolute value of the standardized residual covariance matrix ≤ 0.10, and (2) SRMR $\leq 0.05 \times \bar{R}^2$, where $\bar{R}^2$ is the average communality of the manifest variables in the population. For example, for conditions with medium loadings (.50) on both group and general factors, an exact close fit is achieved if SRMR $= 0.05 \times (0.50^2 + 0.50^2) = 0.025$, and the absolute value of the largest residual is $\leq .10$. The choice of the SRMR was motivated by several reasons. Firstly, the easiness of interpretation of the index. Second, the estimated SRMR is more robust than RMSEA and CFI to different estimation methods, like maximum likelihood and ULS (Xia & Yang, 2019). Finally, the unbiased SRMR is less sensitive than other fit indexes to many of the variables manipulated in the current simulation (i.e., *incidental parameters*; Saris et al., 2009), like the number of items or the number of factors (Fan & Sivo, 2007; Shi et al., 2018; Ximénez et al., 2022). For completeness, we also carried out the simulations without population error to use the results as a baseline for comparison.

4.3 Data generation and analysis

Simulations were run in the R programming language, version 4.1.0 (R Core Team, 2021). A population correlation matrix for each condition was created and stored using the `sim_factor` function from the R package `bifactor`, version 0.1.0 (Jimenez, Abad, Garcia-Garzon, Garrido, & Franco, 2022). Regarding the conditions involving population error, Cudeck and Browne (1992) warned that their method only ensures a global minimum at the intended discrepancy value when the generated error is small enough. Hence, to confirm that close fit was ascertained in each condition, a confirmatory factor analysis using the true model specification was fitted with ULS, and the resulting SRMR was compared with the intended SRMR at a tolerance of 1e-09. Similarly, we also checked whether the estimated parameters were equal to the population parameters. The `sim_factor` function was iterated until a positive definite correlation matrix with error was obtained and satisfied the aforementioned requirements. Table A1 in the appendix displays the average and worst misfit values across every variable level for SRMR, as well as two additional fit indices (CFI and RMSEA), and the maximum absolute residual.

We extracted 50 random samples from a multivariate normal distribution for each correlation matrix using the function `mvrnorm` from the R package `MASS`, version 7.3-57 (Venables & Ripley, 2002). We first estimated the group-factor dimensionality in all these datasets using K1, EKC, PA_{PAF}, PA_{PCA}, and the lowest-level cluster obtained from EGA_{LV}. Second, we performed two oblique factor analyses with ULS, extracting the number of factors suggested by PA_{PCA} and EGA_{LV} and rotating the solution with direct oblimin. Then, we computed the factor scores of each solution using Thurstone's regression method. Finally, for the $\text{PA}_{\text{PCA-FS}}$ method, we performed PA_{PCA} on the factor scores obtained from the first factor solution, whereas for $\text{EGA}_{\text{LV-FS}}$, we used EGA_{LV} on the factor scores obtained from the second factor solution and extracted the highest-level cluster.

We used the function `parallel` from the R package `bifactor` to conduct the methods based on parallel analysis. For all the parallel analysis methods, 100 random datasets were created by within-variable permutation of the empirical dataset to obtain the mean and 95*th* percentile of the eigenvalues under the null model of no latent factors. For the implementation of EGA_{LV}, we used the function `EGA` from the development version of the package `EGAnet`, version 1.1.0 (Golino & Christensen, 2021), available at https://github.com/hfgolino/EGAnet. Importantly, the `EGA` function does not provide the complete hierarchical solution but automatically returns the dimensions that correspond to the highest-level cluster of the hierarchy. Hence, when the LE algorithm determined that the data was not unidimensional, we fed the `cluster_louvain` function from the R package 'igraph", version 1.3.1 (Csardi & Nepusz, 2006), with the estimated network to obtain the complete muti-level organization as estimated by the Louvain algorithm.

Following Garrido et al. (2016) and Golino, Shi, et al. (2020), three indices were calculated to diagnose the accuracy of the methods. The first index is the hit rate (HR) or the proportion of correct dimensionality assessments. While HR reflects each method's accuracy, it does not provide information about the direction of the errors. We thus computed the mean bias error (MBE), conceptualized as the average difference between the estimated dimensionality and the true dimensionality, with positive and negative values reflecting overextraction and underextraction of the true number of factors, respectively. Moreover, as these errors may cancel out in specific conditions, we computed the mean absolute error (MAE), which takes the mean of the absolute error values. Analyses of variance (ANOVA) estimating up to third-order interactions among all the manipulated variables were carried out using the absolute error as the outcome. The partial omega squared (Ω^2) was then used as an effect size to measure each model coefficient's importance. We report all the main effects and only the interactions whose corresponding Ω^2 values were greater than .14 or close to this threshold for at least one method, following Cohen's criterion for a large effect (Cohen, 1988).

All the simulated data, analysis code, and research materials are available at https://osf.io/u7qwj/.

5 Results

Our results suggested that the mean and the 95*th* percentile cut-points behaved similarly across all the levels of the variables in each parallel analysis method. Hence, for simplicity's sake, we will only describe the results of PApca and PApca-fs with the mean value and those of PApaf with the 95*th* percentile[1].

5.1 Recovery of the number of group factors

Overall, EGAlv was the method with the highest hit rate in detecting the number of group factors (HR = .86), closely followed by PApca (HR = .83), and then by PApaf (HR = .64), EKC (HR = .61), and K1 (HR = .60; Table 2). If no population model error existed, PApaf

[1]Overall, the mean value was more accurate than the 95*th* percentile for PApca and PApca-fs whereas the 95*th* percentile was more accurate than the mean value for PApaf.

would have been considered the best method, with an almost perfect hit rate of .98. However, its accuracy was severely impacted when considering model error (HR[MF = close] = .29). In a similar vein, K1 also experimented a strong deterioration under this condition (HR[MF = zero] = .76; HR[MF = close] = .44). On the other hand, the effect of model error on PA_{PCA} was moderate, whereas EKC and EGA_{LV} remained robust to population error.

The number of general factors was a critical variable in our results. Under one general factor, the hit rates of EGA_{LV} and PA_{PCA} were above .95. Whereas increasing the number of general dimensions from one to three decreased the hit rates of EKC and K1 by more than .30 points and those of EGA_{LV} and PA_{PCA} by about .20 points, PA_{PAF} moderately increased its accuracy. However, the accuracy of PA_{PAF} in conditions with three general factors (HR[N.GF = 3] = .65) was still inferior to those of EGA_{LV} (HR[N.GF = 3] = .76) and PA_{PCA}(HR[N.GF = 3] = .74). On the other hand, all the factor retention methods were impaired by the presence of correlations between the general factors, with EGA_{LV} presenting the highest performance in this situation (HR[COR.GF = .30] = .84).

However, EGA_{LV} did not always perform best. While it attained almost perfect accuracy in simple structures (HR[CROSS.GRF = 0] = .99), it showed drops of .10 (HR = .89) and .29 points (HR = .70) when the size of the cross-loadings increased to .15 and .30, respectively. On the contrary, PA_{PCA} and EKC were only moderately affected by the presence of high cross-loadings, with the former attaining the best average performance across high cross-loadings conditions (HR[CROSS.GRF = .30] = .79). Conversely, PA_{PAF} and K1 were not affected by item complexity, but their performances were still inferior to those of EGA_{LV} and PA_{PCA} .

Increasing the number of group factors per general factor negatively affected all the methods. EGA_{LV} and PA_{PAF} were only moderately affected, with the former retaining the highest accuracy across all the levels. However, K1, EKC, and PA_{PCA} were more affected by the increase in the number of group factors from four to six, showing declines of .16, .14, and .10 points in accuracy, respectively. On the other hand, increasing the number of variables per group factor also increased the accuracy of all the methods but K1 and PA_{PAF}. K1 was the most accurate method across conditions with four variables per group factor (HR[VAR.GRF = 4] = .90), but the worst across conditions with eight and ten variables (HR[VAR.GRF = 10] = .34). Conversely, the accuracy of EKC was severely impaired in conditions where four variables loaded on the group factors instead of on six (HR[VAR.GRF = 4] = .24; HR[VAR.GRF = 6] = .60). PA_{PCA} also benefited by switching from four to six variables per group factor (HR[VAR.GRF = 4] = .60; HR[VAR.GRF = 6] = .89), but further increases in the number of variables per group factor did not produce substantial gains in accuracy[2]. Concerning EGA_{LV}, it obtained the best hit rate in conditions with the maximum number of variables per group factor (HR[VAR.GRF = 10] = .96).

We further identified three results of interest. When switching from medium to low loadings on the group factors, K1, EKC, PA_{PCA}, and EGA_{LV} were negatively impacted, with respective hit rate drops of .16, .27, .17, and .13 points. Again, EGA_{LV} was the best method across the most unfavorable condition (e.g., HR[LOAD.GRF = low] = .80). Secondly, concerning the loadings on the general factors, lower loadings were moderately associated with higher hit rates for EKC and PA_{PCA} (with absolute increases of .07 and .09 points, respectively) but negatively impacted K1 with a drop of .16 points. EGA_{LV} remained unaffected to the magnitude of the loadings on the general factors, whereas PA_{PAF} was robust to the magnitude of the general and group factor loadings. Lastly, the sample size was positively related to the hit rate of all the factor retention methods, with PA_{PAF} being again the exemption. While PA_{PAF} presented a good average performance across small sample sizes (HR[N = 500] = .80), it drastically underperformed as the sample size increased (e.g., HR[N = 5000] = .51). Interestingly, sample size had very little influence on EGA_{LV}, and for conditions with a sample size of 2000 or greater, PA_{PCA} slightly outperformed EGA_{LV} with a hit rate about .90. K1 and EKC benefited from increased sample sizes but only achieved an overall hit rate over .80 across conditions with a sample of size 5000.

[2] We verified that this lack of improvement for PA_{PCA} was due to the presence of population error. Removing the conditions with population error yielded a clearer increasing monotonic relationship between the hit rate and VAR.GRF.

The results for the mean bias error (MBE; Table 3) revealed that, following the HR results, EGA_{LV} and PA_{PCA} were the least biased methods. EGA_{LV}, PA_{PCA}, and EKC underestimated the number of factors, with overall MBEs of -0.29, -0.44, and -1.86, respectively. EGA_{LV} underextracted the most in conditions involving few variables per group factor (MBE[VAR.GRF = 4] = -0.76) and high cross-loadings (MBE[CROSS.GRF = .30] = -0.75). The worst performance of PA_{PCA} was observed under weakly defined group factors (MBE[VAR.GRF = 4] = -1.46; MBE[LOAD.GRF = low] = -0.82) and low sample size (MBE[N = 500] = -1.14). EKC was severely biased in these same conditions (e.g., MBE[VAR.GRF = 4] = -4.32; MBE[LOAD.GRF = low] = -2.78), only obtaining acceptable performance when either the sample size or the number of variables per group factor were high. Contrary to the underestimation of the previous methods, K1 and PA_{PAF} overextracted across all the variable levels with the exemption of PA_{PAF} in conditions with no population error, in which it was unbiased. Their overall MBEs were 2.07 and 1.58, respectively, with K1 being particularly prone to overextraction in situations involving small sample size (MBE[N = 500] = 4.84), large factor structures (MBE[VAR.GRF = 10] = 4.64; MBE[N.GF = 3] = 3.45; MBE[NUM.GRF = 6] = 2.96), and low loadings on both the general and group factors (MBE[LOAD.GF = low] = 2.95; MBE[LOAD.GRF = low] = 2.95). K1 only showed an acceptable performance for the conditions involving the maximum sample size and the minimum number of variables per group factor. The performance of PA_{PAF} was particularly hindered in large sample size conditions (MBE[N = 5000] = 3.75), population structures with population error (MBE[MF = close] = 3.17), and correlated general factors (MBE[COR.GF = .30] = 2.51). Despite PA_{PAF} not being influenced by the number of variables per group factor in terms of accuracy, the MBE indicated that it overextracted more factors the more variables defined a group factor. In the end, PA_{PAF} only showed an acceptable overall performance for population structures without error and across conditions with the minimum sample size.

Because the estimation biases may cancel out when computing marginal means, we further assessed the precision of the factor retention methods with the MAE (Table 4). However, the MAE followed a similar pattern to the MBE across all the manipulated levels and will not be further discussed.

As the overall performances of K1, EKC, and PA_{PAF} were much worse than those of PA_{PCA} and EGA_{LV}, in Table 5, we only show the Ω^2 effect sizes obtained for PA_{PCA} and EGA_{LV} from the analysis of variance[3]. PA_{PCA} was most sensitive to VAR.GRF, a variable also involved in all the large two-way and three-way interactions. These interactions showed that the effect of other variables (LOAD.GF, LOAD.GRF, N, and N.GF) was smaller as the number of variables per group factor increased. Lower loadings on the group factors were very detrimental when the group factors were defined by fewer variables, especially in smaller samples (Figure 3(a); Ω^2[VAR.GRF $\times$ N $\times$ LOAD.GRF] = .22). Similarly, having more general factors was increasingly deleterious when fewer variables loaded on the group factors, particularly when the sample size was smaller (Figure 3(b); Ω^2[VAR.GRF $\times$ N $\times$ N.GF] = .18). Noteworthy, for samples of size 1000 or larger and at least six indicators per group factor, the negative effect of having lower loadings on the group factors and more general factors was small. Another three-way interaction indicated that PA_{PCA} tended to underperform more with lower loadings on the group factors when fewer variables defined them and when there were more general factors (Figure 4; Ω^2[VAR.GRF $\times$ N.GF $\times$ LOAD.GRF] = .16). In other words, with an increasing number of general factors, more indicators per group factor might be needed if their quality is low. Finally, an interaction indicated that higher loadings on the general factors were more detrimental when the group factors were defined by only a few items (Figure 5; Ω^2[VAR.GRF $\times$ LOAD.GF] = .19). That is, better-defined group factors counterbalanced the effect induced by the presence of stronger general factors (e.g., higher correlations among the variables that loaded on the same general factor but different group factors).

Concerning EGA_{LV}, the results of the ANOVA revealed that it was sensitive to the number of variables per group factor, the number of general factors, and the presence of cross-

[3]Readers interested in the most relevant effect sizes found for K1, EKC, and PA_{PAF} can find them in the Table A2 from the appendix.

loadings among the group factors. All the effects produced by these variables were smaller on EGA_{LV} than on PA_{PCA}, except those involving cross-loadings. When there were no cross-loadings, EGA_{LV} remained robust to weakly defined group factors (i.e., few variables per group factor with low loadings), and larger factor structures. Small cross-loadings started to become detrimental only in structures with three general factors or low loadings on the group factors if the number of variables per group factor was eight or smaller. However, the effect of high cross-loadings was very detrimental when the group factors had fewer variables in structures with more than one general factor (Figure 6(a); Ω^2[VAR.GRF $\times$ CROSS.GRF $\times$ N.GF] = .22) or with lower loadings on the group factors (Figure 6(b); Ω^2[VAR.GRF $\times$ CROSS.GRF $\times$ N.GF] = .13). Such detrimental effect of cross-loadings, in interaction with the aforementioned variables, was small whenever eight or more variables defined each group factor.

5.2 Recovery of the number of general factors

Despite the good performance of the lowest-level cluster of EGA_{LV} in identifying the number of group factors, it only identified a higher layer of clusters in 42% of the simulated datasets. Even in these cases, it often provided a wrong estimation of the number of general factors, with an overall hit rate of .24. Therefore, we did not seek to analyze this method in further analyses. Similarly, K1, EKC, and PA_{PAF} were inaccurate for detecting the number of group factors in many situations, so they were not further considered, as explained before. In contrast, the estimation of the number of general factors was extraordinarily accurate using either $\text{PA}_{\text{PCA-FS}}$ or $\text{EGA}_{\text{LV-FS}}$. These methods presented hit rates close to one and mean absolute errors close to zero across all the variable levels (Tables 2 and 4). The minimum marginal hit rates and maximum marginal mean absolute errors for $\text{PA}_{\text{PCA-FS}}$ happened in the conditions with few variables per group factor (HR = .97, MAE = 0.04, VAR.GRF = 4) and small sample size (HR = .97, MAE = 0.03, N = 500). On the other hand, $\text{EGA}_{\text{LV-FS}}$ had an almost perfect performance across all the variable levels. Interestingly, none of the estimated

Ω^2 effect sizes for either method were high (Table 5). For $\mathrm{PA}_{\text{PCA-FS}}$, the maximum Ω^2 value associated with a main effect was 0.03, and for $\mathrm{EGA}_{\text{LV-FS}}$, 0.01.

6 Discussion

Dimensionality assessment is one of the most important decisions that researchers face in test development and validation. It is well known that wrong dimensionality assessments can severely bias item parameter estimates and undermine the validity of test scores (Fava & Velicer, 1992, 1996). Unfortunately, theory is not always enough to ascertain the number of factors underlying a dataset, and factor retention methods become necessary. Today, there is little information on how to assess the dimensionality of structures with factors subsumed into broader, higher-order factors, like those encountered in intelligence, personality, and psychopathology. While many bi-factor and generalized bi-factor methods have been developed recently to estimate large and complex structures that account for the presence of general factors (Abad et al., 2017; Cai, 2010; Garcia-Garzon et al., 2019, 2020; Jennrich & Bentler, 2011; Jimenez, Abad, Garcia-Garzon, & Garrido, 2022; Nájera et al., 2021), we still lack evidence-based recommendations on how to assess the dimensionality of this kind of structures. This is a crucial limitation because all of these methods assume that the number of group and general factors are known.

Hence, in this study, we investigated for the first time the performance of some classical and recent factor retention methods to uncover the number of group and general factors in bi-factor structures up to three general factors. Overall, we found that EGA_{LV} was the most accurate, precise, and robust method for estimating the number of group factors, followed by PA_{PCA}, which was sensitive to various conditions, namely the number of variables per group factor, sample size, and loadings on the group and general factors. These results align with previous research showing that PA_{PCA} underestimates the number of factors in conditions involving small samples and large factor structures with weakly defined group factors (Braeken & Assen, 2017; Garrido et al., 2013; Yang & Xia, 2015). Notwithstanding, the performance of PA_{PCA} was very high whenever the sample size was above 1000, and the number of variables per group factor was six or higher. Our findings also agree with previous results in which EGA was highly robust to unfavorable conditions, albeit using the Walktrap clustering algorithm instead of Louvain (Cosemans et al., 2021; Golino & Epskamp, 2017; Golino, Shi, et al., 2020). The other tested factor retention methods, K1, EKC, and PA_{PAF}, did not perform well in estimating the number of group factors and were not further examined.

Interestingly, sample size and model misfit had little influence on EGA_{LV}. A possible explanation for the latter finding is that the GLASSO penalization shrinks towards zero small partial correlations that appear due to trivial common variance attributable to population error. However, the performance of EGA_{LV} was not perfect. It was sensitive to high cross-loadings, particularly in factor structures with more than one general factor and weakly defined group factors. This sensitivity of EGA_{LV} to high cross-loadings could be due to the fact that the Louvain algorithm does not allow overlapping clusters (Blanken et al., 2018; Christensen et al., 2020). In other words, items cannot be simultaneously classified in more than one cluster, which increases the probability of incorrect placements if cross-loadings exist. Nonetheless, this problem could be ameliorated in the future by replacing Louvain with a clustering algorithm that does not have this limitation or by using network loadings to verify the structure suggested by the algorithm (Christensen & Golino, 2021a).

Within the parallel analysis methods, many researchers have suggested that PA_{PAF} is more suitable than PA_{PCA} for correlated psychological data, both theoretically and empirically (Crawford et al., 2010; Green et al., 2012; Keith et al., 2016). Particularly, Crawford et al. (2010) found that PA_{PAF} performed better than PA_{PCA} under multiple correlated factors, second-order general factors, and bi-factor models. However, they did not consider the role of population error in their simulations. As revealed in our results and in other studies such as Lim and Janhg (2019) and Xia (2021), the accuracy of PA_{PAF} greatly diminishes in the presence of trivial population misfit and only outperforms other methods if, and only if, no population error exists. Unfortunately, some sort of population misfit is always expected to exist in applied settings. Moreover, PA_{PAF} tended to overextract with higher sample sizes and an increasing number of variables per group factor. Therefore, we consider that PA_{PAF} is inappropriate for evaluating the dimensionality of bi-factor and generalized bi-factor structures. Contrary, PA_{PCA} was only moderately affected by the presence of close misfit, a result that is also consistent with previous research (Lim & Jahng, 2019; Xia, 2021). On the other hand, using either the mean value or the 95th percentile as the cut-off for computing the reference eigenvalues did not result in a practical difference for PA_{PCA} .

Overall, EKC had a modest performance and was less robust than EGA_{LV} and PA_{PCA} to most of the manipulated variables (Table A2, appendix). To our surprise, it was comparable to that of K1. However, whereas EKC displayed an increased performance the more variables defined the group factors, K1 exceedingly overextracted factors. Additionally, K1 was very sensitive to population error, while EKC was not influenced by it. Thus, our results agree with several decades of simulation research in that K1 should never be used for dimensionality assessment, especially in large factor structures like the ones often encountered in bi-factor applications.

Regarding the estimation of the general factors, we found that when EGA_{LV} estimated more than one layer of clusters, the number of factors suggested by the highest-level cluster was mostly inaccurate. On the contrary, $EGA_{LV\text{-}FS}$ and $PA_{PCA\text{-}FS}$ had almost perfect accuracy across all the conditions and were robust to all the manipulated variables. This means that the number of general factors could be estimated accurately even when EGA_{LV} and PA_{PCA} failed to determine the correct number of group factors. However, we do not recommend applying these methods blindly. These second-order methods should only be considered when the correlations between the factor scores are not trivially small. In other words, we recommend inspecting the first-order factor correlation matrix before interpreting the estimates provided by $EGA_{LV\text{-}FS}$ and $PA_{PCA\text{-}FS}$. Otherwise, we would be at risk of inferring the presence of general factors when there is no more variance to explain beyond the one accounted for the first-order factors.

An advantage of our second-order proposals over Goldberg's Bass-Ackwards method is that they are based on a bottom-up approach. We first focus on estimating the number of lower-order factors and then proceed with the higher-order ones. This way, we are able to identify the nuances that make up the more general traits, encouraging the analysis of item content and domain's breadth (Condon et al., 2020; Mõttus et al., 2020). We also remark that $\text{EGA}_{\text{LV-FS}}$ is somewhat similar to the second-order method proposed by Golino, Jotheeswaran, et al. (2020). The main differences between our and their approach are that we used the lowest-level cluster provided by the Louvain algorithm instead of Walktrap and analyzed the correlation matrix between the factor scores instead of the correlation matrix between the rotated factors, which does not require computing the factor scores. Future simulation studies may consider including the method of Golino, Jotheeswaran, et al. (2020) to check whether it performs as well as $\text{EGA}_{\text{LV-FS}}$.

This simulation study tried to emulate real data with conditions involving population misfit and cross-loadings, but it has some limitations: first, we only generated continuous data from multivariate normal distributions. With categorical data, polychoric correlation matrices, and skewed distributions, the performance of all the methods should deteriorate, and the extent to which this would happen is unknown. Second, we only generated factor structures up to three general factors, whereas some cases of psychological data may contain more. This limitation was due to the fact that controlling population misfit in conditions involving more than three general factors is a difficult task, as larger factor structures produce correlation matrices closer to nonpositiveness. Forthcoming work will be needed to solve these technical issues inherent to generalized bi-factor structures. Notwithstanding, the current simulation is the first one that systematically investigates the dimensionality assessment of factor structures with a varying number of general factors, and it is a good first step toward developing tools for factor retention in fields like intelligence, personality, and psychopathology, where the statistical models usually display a hierarchical configuration.

In conclusion, we aimed to provide applied researchers with accurate methods that can help them to uncover hierarchical structures in their data, and our results suggest that parallel analysis with principal component analysis and exploratory graph analysis with the Louvain algorithm, when applied to items and then to the first-order factor scores, offer a good recovery of the dimensionality of the hierarchical structure. As different variables impact these two methods, researchers may use them in tandem or according to the known or plausible characteristics of their data. Noteworthy, EGA_{LV} not only was the best method in terms of accuracy, precision, and robustness for the conditions most likely to be encountered in practice, but also provides a classification of items into factors, offering a richer dimensionality assessment that can be easily compared with the theoretical expectations of the factor structure. Furthermore, the stability of the EGA_{LV} and $\text{EGA}_{\text{LV-FS}}$ latent solutions can be readily ascertained using bootstrap procedures currently available (Christensen & Golino, 2021b). Thus, we highlight the particular usefulness of EGA_{LV} and $\text{EGA}_{\text{LV-FS}}$ for assessing bi-factor and generalized bi-factor structures. Finally, much more attention should be considered to the number of group factors, as the second-order methods depend on this quantity, and they are harder to estimate than the number of general factors.

References

Abad, F. J., Garcia-Garzon, E., Garrido, L. E., & Barrada, J. R. (2017). Iteration of partially specified target matrices: Application to the bi-factor case. *Multivariate Behavioral Research*, *52*(4), 416–429. https://doi.org/10.1080/00273171.2017.1301244

Abad, F. J., Sorrel, M. A., Garcia, L. F., & Aluja, A. (2018). Modeling general, specific, and method variance in personality measures: Results for ZKA-PQ and NEO-PI-R. *Assessment*, *25*(8), 959–977. https://doi.org/10.1177/1073191116667547

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468–491. https://doi.org/10.1037/met0000200

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-Factor vs. Higher-order models. *Journal of Intelligence*, *3*(4), 121–136. https://doi.org/10.3390/jintelligence3040121

Blanken, T. F., Deserno, M. K., Dalege, J., Borsboom, D., Blanken, P., Kerkhof, G. A., & Cramer, A. O. J. (2018). The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks. *Scientific Reports*, *8*(1), 5854. https://doi.org/10.1038/s41598-018-24224-2

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bork, R. van, Rhemtulla, M., Waldorp, L. J., Kruis, J., Rezvanifar, S., & Borsboom, D. (2021). Latent variable models and networks: Statistical equivalence and testability. *Multivariate Behavioral Research*, *56*(2), 175–198. https://doi.org/10.1080/00273171.2019.1672515

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry*, *88*(1), 18–27. https://doi.org/10.1016/j.biopsych.2020.01.013

Braeken, J., & Assen, M. A. L. M. van. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, *27*(4), 509–540. https://doi.org/10.1207/s15327906mbr2704_2

Cai, L. (2010). A Two-Tier Full-Information Item Factor Analysis Model with Applications. *Psychometrika*, *75*(4), 581–612. https://doi.org/10.1007/s11336-010-9178-0

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Christensen, A. P., Garrido, L. E., & Golino, H. (2020). *Comparing community detection algorithms in psychological data: A Monte Carlo simulation*. PsyArXiv. https://doi.org/10.31234/osf.io/hz89e

Christensen, A. P., & Golino, H. (2021a). On the equivalency of factor and network loadings. *Behavior Research Methods*, *53*(4), 1563–1580. https://doi.org/10.3758/s13428-020-01500-6

Christensen, A. P., & Golino, H. (2021b). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. *Psych*, *3*(3), 479–500. https://doi.org/10.3390/psych3030032

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, *103*(2), 276–279. https://doi.org/10.1037/0033-2909.103.2.276

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge Academic. https://www.routledge.com/Statistical-Power-Analysis-for-the-Behavioral-Sciences/Cohen/p/book/9780805802832

Condon, D. M., Wood, D., Mõttus, R., Booth, T., Costantini, G., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Ziegler, M., & Zimmermann, J. (2020). Bottom up construction of a personality taxonomy. *European Journal of Psychological Assessment*, *36*(6), 923–934. https://doi.org/10.1027/1015-5759/a000626

Cosemans, T., Rosseel, Y., & Gelper, S. (2021). Exploratory Graph Analysis for factor retention: Simulation results for continuous and binary data. *Educational and Psychological Measurement*, 00131644211059089. https://doi.org/10.1177/00131644211059089

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, *70*(6), 885–901. https://doi.org/10.1177/0013164410379332

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. https://igraph.org

Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika*, *57*(3), 357–369. https://doi.org/10.1007/BF02295424

Epskamp, S., & Fried, E. (2017). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*. https://doi.org/10.1037/met0000167

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272. https://doi.org/10.1037/1082-989X.4.3.272

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529. https://doi.org/10.1080/00273170701382864

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, *27*(3), 387–415. https://doi.org/10.1207/s15327906mbr2703_5

Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analyses. *Educational and Psychological Measurement*, *56*(6), 907–929. https://doi.org/10.1177/0013164496056006001

Friborg, O., Hjemdal, O., Martinussen, M., & Rosenvinge, J. H. (2009). Empirical support for resilience as more than the counterpart and absence of vulnerability and symptoms of mental disorder. *Journal of Individual Differences*, *30*(3), 138–151. https://doi.org/10.1027/1614-0001.30.3.138

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

Garcia-Garzon, E., Abad, F. J., & Garrido, L. E. (2019). Improving bi-factor exploratory modeling. *Methodology*, *15*(2), 45–55. https://doi.org/10.1027/1614-2241/a000163

Garcia-Garzon, E., Nieto, M. D., Garrido, L. E., & Abad, F. J. (2020). Bi-factor exploratory structural equation modeling done right: Using the SLiDapp application. *Psicothema*, *32.4*, 607–614. https://doi.org/10.7334/psicothema2020.179

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

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

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*(3), 377–393. https://doi.org/10.1177/0013164495055003002

Goldberg, L. R. (2006). Doing it all Bass-Ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality*, *40*(4), 347–358. https://doi.org/10.1016/j.jrp.2006.01.001

Golino, H., & Christensen, A. P. (2021). *EGAnet: Exploratory Graph Analysis – A framework for estimating the number of dimensions in multivariate data using network psychometrics* [Manual].

Golino, H., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis. *Intelligence*, *62*, 54–70. https://doi.org/10.1016/j.intell.2017.02.007

Golino, H., & Epskamp, S. (2017). Exploratory Graph Analysis: A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, *12*(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Jotheeswaran, A., Sadana, R., Teles, M., Christensen, A., & Boker, S. (2020). *Investigating the broad domains of intrinsic capacity, functional ability and environment: An exploratory graph analysis approach for improving analytical methodologies for measuring healthy aging.* https://doi.org/10.31234/osf.io/hj5mc

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

Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, *40*(7), 3510–3521. https://doi.org/10.1007/s12144-019-00300-2

Green, S. B., Levy, R., Thompson, M. S., Lu, M., & Lo, W.-J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel nnalysis. *Educational and Psychological Measurement*, *72*(3), 357–374. https://doi.org/10.1177/0013164411422252

Green, S. B., Redell, N., Thompson, M. S., & Levy, R. (2016). Accuracy of revised and traditional parallel analyses for assessing dimensionality with binary data. *Educational and Psychological Measurement*, *76*(1), 5–21. https://doi.org/10.1177/0013164415581898

Green, S. B., Thompson, M. S., Levy, R., & Lo, W.-J. (2015). Type I and Type II error rates and overall accuracy of the revised parallel analysis method for determining the number of factors. *Educational and Psychological Measurement*, *75*(3), 428–457. https://doi.org/10.1177/0013164414546566

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149–161. https://doi.org/10.1007/BF02289162

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

Humphreys, L. G., & Ilgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement*, *29*(3), 571–578. https://doi.org/10.1177/001316446902900303

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory Bi-Factor Analysis. *Psychometrika*, *76*(4), 537–549. https://doi.org/10.1007/s11336-011-9218-4

Jimenez, M., Abad, F. J., Garcia-Garzon, E., & Garrido, L. E. (2022). *Dimensionality assessment in generalized bi-factor structures.* https://osf.io/u7qwj/

Jimenez, M., Abad, F. J., Garcia-Garzon, E., Garrido, L. E., & Franco, V. R. (2022). *Bifactor: Exploratory factor, bi-factor, and generalized bi-factor modeling* [Manual]. https://github.com/Marcosjnez/bifactor

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Keith, T. Z., Caemmerer, J. M., & Reynolds, M. R. (2016). Comparison of methods for factor extraction for cognitive test-like data: Which overfactor, which underfactor? *Intelligence*, *54*, 37–54. https://doi.org/10.1016/j.intell.2015.11.003

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Levy, R., Xia, Y., & Green, S. B. (2021). Incorporating uncertainty into parallel analysis for choosing the number of factors via bayesian methods. *Educational and Psychological Measurement*, *81*(3), 466–490. https://doi.org/10.1177/0013164420942806

Li, Y., Wen, Z., Hau, K.-T., Yuan, K.-H., & Peng, Y. (2020). Effects of cross-loadings on determining the number of factors to retain. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(6), 841–863. https://doi.org/10.1080/10705511.2020.1745075

Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods*, *24*(4), 452–467. https://doi.org/10.1037/met0000230

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research*, *38*(1), 113–139. https://doi.org/10.1207/S15327906MBR3801_5

Mansolf, M., & Reise, S. P. (2016). Exploratory Bifactor Analysis: The Schmid-Leiman Orthogonalization and Jennrich-Bentler Analytic Rotations. *Multivariate Behavioral Research*, *51*(5), 698–717. https://doi.org/10.1080/00273171.2016.1215898

Marčenko, V. A., & Pastur, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Math USSR Sb*, *1*, 457–483.

Milfont, T. L., & Duckitt, J. (2004). The structure of environmental attitudes: A first- and second-order confirmatory factor analysis. *Journal of Environmental Psychology*, *24*(3), 289–303. https://doi.org/10.1016/j.jenvp.2004.09.001

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big Few traits. *European Journal of Personality*, *34*(6), 1175–1201. https://doi.org/10.1002/per.2311

Nájera, P., Abad, F. J., & Sorrel, M. A. (2021). Determining the number of attributes in cognitive diagnosis modeling. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.614470

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

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10*(2), 191–218. https://eudml.org/doc/55419

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

Reise, S. P. (2012). The Rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Ruscio, J., & Roche, B. (2011). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*, 282–292. https://doi.org/10.1037/a0025697

Saris, W. E., Satorra, A., & Veld, W. M. van der. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582. https://doi.org/10.1080/10705510903203433

Shi, D., Maydeu-Olivares, A., & DiStefano, C. (2018). The relationship between the standardized root mean square residual and model misspecification in factor analysis models. *Multivariate Behavioral Research*, *53*(5), 676–694. https://doi.org/10.1080/00273171.2018.1476221

Tian, C., & Liu, Y. (2021). A Rotation Criterion That Encourages a Hierarchical Factor Structure. In M. Wiberg, D. Molenaar, J. González, U. Böckenholt, & J.-S. Kim (Eds.), *Quantitative Psychology* (pp. 1–8). Springer International Publishing. https://doi.org/10.1007/978-3-030-74772-5_1

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (4th ed.). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Xia, Y. (2021). Determining the number of factors when population models can be closely approximated by parsimonious models. *Educational and Psychological Measurement*, *81*(6), 1143–1171. https://doi.org/10.1177/0013164421992836

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–

428. https://doi.org/10.3758/s13428-018-1055-2

Ximénez, C., Maydeu-Olivares, A., Shi, D., & Revuelta, J. (2022). Assessing cutoff values of SEM fit indices: Advantages of the unbiased SRMR index and its cutoff criterion based on communality. *Structural Equation Modeling: A Multidisciplinary Journal*, *0*(0), 1–13. https://doi.org/10.1080/10705511.2021.1992596

Yang, Y., & Xia, Y. (2015). On the number of factors to retain in exploratory factor analysis for ordered categorical data. *Behavior Research Methods*, *47*(3), 756–772. https://doi.org/10.3758/s13428-014-0499-2

Yeomans, K. A., & Golder, P. A. (1982). The Guttman-Kaiser criterion as a predictor of the number of common factors. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *31*(3), 221–229. https://doi.org/10.2307/2987988

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432–442. https://doi.org/10.1037/0033-2909.99.3.432

Tables

Table 1. Simulated loadings for a condition with one general factor, four group factors and medium loadings on both the general and group factors. When cross-loadings (underlined) were included, small values were subtracted from the loadings on the general and group factors to maintain the original communality ($\mathbf{h^2}$).

Item	Simple structure						Cross-loadings					
	G	S1	S2	S3	S4	$\mathbf{h^2}$	G	S1	S2	S3	S4	$\mathbf{h^2}$
1	.45	.60				.57	.40	.56			.30	.57
2	.47	.53				.51	.47	.53				.51
3	.51	.47				.48	.51	.47				.48
4	.58	.40				.50	.58	.40				.50
5	.44		.60			.55	.39	.30	.56			.55
6	.58		.53			.62	.58		.53			.62
7	.59		.47			.56	.59		.47			.56
8	.53		.40			.44	.53		.40			.44
9	.53			.60		.64	.48		.30	.56		.64
10	.41			.53		.45	.41			.53		.45
11	.44			.47		.41	.44			.47		.41
12	.44			.40		.35	.44			.40		.35
13	.54				.60	.65	.49			.30	.56	.65
14	.48				.53	.51	.48				.53	.51
15	.55				.47	.52	.55				.01	.52
16	.50				.40	.41	.50				.40	.41
Avg.						.51						.51

Table 2. Marginal hit rates across each variable level for each factor retention method.

Variable	Group factors							General factors			
	Kaiser		PA_{PAF}		PA_{PCA}		EGA_{LV}	$PA_{PCA\text{-}FS}$		EGA_{LV}	$EGA_{LV\text{-}FS}$
	K1	EKC	mean	95*th*	mean	95*th*		mean	95*th*		
MF											
zero	.76	.62	.98	.98	.86	.84	.87	.99	.98	.10	1.00
close	.44	.61	.21	.29	.81	.80	.86	.99	.99	.09	1.00
N											
500	.33	.28	.72	.79	.68	.64	.84	.97	.95	.06	1.00
1000	.55	.55	.63	.69	.85	.83	.86	.99	.99	.09	1.00
2000	.72	.75	.53	.57	.90	.89	.87	1.00	1.00	.11	1.00
5000	.81	.87	.50	.51	.91	.91	.89	1.00	1.00	.13	1.00
N.GF											
1	.78	.82	.52	.58	.95	.94	.98	1.00	1.00	.00	1.00
2	.61	.64	.60	.66	.87	.85	.91	.99	.99	.02	1.00
3	.49	.49	.62	.65	.74	.73	.76	.98	.98	.22	.99
COR.GF											
0	.67	.65	.64	.69	.87	.86	.88	.99	.99	.09	1.00
.30	.50	.55	.52	.56	.77	.76	.84	.99	.98	.10	1.00
VAR.GRF											
4	.90	.24	.59	.67	.60	.56	.75	.97	.94	.25	.99
6	.68	.60	.59	.64	.89	.87	.83	1.00	1.00	.13	1.00
8	.48	.77	.59	.63	.93	.92	.92	1.00	1.00	.01	1.00
10	.34	.84	.60	.62	.92	.92	.95	1.00	1.00	.00	1.00
NUM.GRF											
4	.68	.68	.61	.67	.88	.87	.90	.99	.98	.10	1.00
5	.60	.61	.59	.64	.84	.82	.86	.99	.99	.10	1.00
6	.52	.54	.57	.61	.78	.76	.83	.99	.99	.10	1.00
CROSS.GRF											
0	.61	.65	.60	.65	.87	.86	.99	.99	.98	.07	1.00
.15	.60	.62	.60	.64	.84	.82	.89	.99	.98	.08	1.00
.30	.59	.57	.58	.63	.79	.78	.70	.99	.99	.14	.99
LOAD.GRF											
low	.52	.48	.59	.64	.75	.72	.80	.98	.97	.18	.99
medium	.68	.75	.60	.63	.92	.91	.93	1.00	1.00	.02	1.00
LOAD.GF											
low	.52	.65	.60	.66	.88	.87	.86	1.00	1.00	.07	1.00
medium	.68	.58	.59	.62	.79	.77	.87	.98	.97	.13	1.00
Total	.60	.61	.59	.64	.83	.82	.86	.99	.99	.10	1.00

Note. K1 = Kaiser eigenvalue greater-than-one criterion; EKC = Empirical Kaiser Criterion; PA_{PAF} = Parallel analysis with principal axis factoring; PA_{PCA} = Parallel analysis with principal components; $PA_{PCA\text{-}FS}$ = Parallel analysis with principal components on the first-order factor scores; EGA = Exploratory Graph Analysis; EGA_{LV} = EGA with Louvain; $EGA_{LV\text{-}FS}$ = EGA with Louvain on the first-order factor scores; MF = population misfit; N = sample size; N.GF = number of general factors; COR.GF = correlation between general factors; VAR.GRF = number of indicators per group factor; NUM.GRF = number of group factors per general factor; CROSS.GRF = cross-loadings in the group factors; LOAD.GRF = loadings on the group factors; LOAD.GF = loadings on the general factors.

Table 3. Mean bias error (MBE) across each variable level for each factor retention method.

Variable	Group factors							General factors			
	Kaiser		PAPAF		PAPCA		EGALV	PAPCA-FS		EGALV	EGALV-FS
	K1	EKC	mean	95*th*	mean	95*th*		mean	95*th*		
MF											
zero	1.56	-1.88	0.01	-0.01	-0.48	-0.57	-0.30	-0.01	-0.02	0.64	0.00
close	2.58	-1.84	3.83	3.17	-0.40	-0.49	-0.29	-0.01	-0.02	0.70	0.00
N											
500	4.84	-4.04	0.40	0.16	-1.13	-1.37	-0.26	-0.03	-0.07	1.48	0.00
1000	2.25	-2.11	1.00	0.68	-0.44	-0.54	-0.31	0.00	-0.01	0.86	0.00
2000	0.85	-0.97	2.12	1.72	-0.16	-0.20	-0.31	0.00	0.00	0.39	0.00
5000	0.33	-0.32	4.17	3.76	-0.01	-0.02	-0.30	0.00	0.00	-0.05	-0.01
N.GF											
1	0.41	-0.34	2.43	1.99	-0.07	-0.10	0.02	0.00	0.00	-0.88	0.00
2	1.51	-1.40	1.73	1.41	-0.31	-0.39	-0.14	0.00	-0.01	0.12	0.00
3	3.46	-3.08	1.86	1.54	-0.75	-0.89	-0.61	-0.02	-0.04	2.00	-0.01
COR.GF											
0	1.61	-1.61	1.23	0.96	-0.40	-0.48	-0.26	-0.01	-0.02	0.40	0.00
.30	2.76	-2.23	2.96	2.51	-0.49	-0.61	-0.35	-0.01	-0.02	1.08	0.00
VAR.GRF											
4	0.09	-4.32	0.88	0.59	-1.46	-1.73	-0.76	-0.04	-0.08	1.41	-0.01
6	0.99	-1.84	1.65	1.32	-0.29	-0.36	-0.28	0.00	0.00	1.20	0.00
8	2.55	-0.86	2.31	1.94	-0.04	-0.07	-0.10	0.00	0.00	0.40	0.00
10	4.64	-0.42	2.85	2.47	0.05	0.03	-0.04	0.00	0.00	-0.33	0.00
NUM.GRF											
4	1.23	-1.14	1.60	1.28	-0.24	-0.31	-0.17	-0.01	-0.02	-0.26	-0.01
5	2.02	-1.82	1.92	1.58	-0.42	-0.51	-0.29	-0.01	-0.02	0.61	0.00
6	2.96	-2.63	2.25	1.88	-0.65	-0.77	-0.42	-0.01	-0.02	1.67	0.00
CROSS.GRF											
0	2.06	-1.72	2.00	1.66	-0.33	-0.42	0.00	-0.01	-0.02	0.20	0.00
.15	2.06	-1.89	1.95	1.62	-0.43	-0.53	-0.14	-0.01	-0.03	0.62	0.00
.30	2.09	-1.97	1.81	1.47	-0.55	-0.65	-0.75	0.00	-0.01	1.19	-0.01
LOAD.GRF											
low	2.95	-2.78	1.90	1.52	-0.82	-0.98	-0.48	-0.02	-0.04	1.32	-0.01
medium	1.19	-0.94	1.94	1.64	-0.05	-0.08	-0.11	0.00	0.00	0.03	0.00
LOAD.GF											
low	2.97	-1.58	1.68	1.34	-0.24	-0.31	-0.28	0.00	0.00	0.80	0.00
medium	1.17	-2.14	2.16	1.82	-0.63	-0.75	-0.31	-0.02	-0.04	0.55	-0.01
Total	2.07	-1.86	1.92	1.58	-0.44	-0.53	-0.29	-0.01	-0.02	0.67	0.00

Note. K1 = Kaiser eigenvalue greater-than-one criterion; EKC = Empirical Kaiser Criterion; PAPAF = Parallel analysis with principal axis factoring; PAPCA = Parallel analysis with principal components; PAPCA-FS = Parallel analysis with principal components on the first-order factor scores; EGA = Exploratory Graph Analysis; EGALV = EGA with Louvain; EGALV-FS = EGA with Louvain on the first-order factor scores; MF = population misfit; N = sample size; N.GF = number of general factors; COR.GF = correlation between general factors; VAR.GRF = number of indicators per group factor; NUM.GRF = number of group factors per general factor; CROSS.GRF = cross-loadings in the group factors; LOAD.GRF = loadings on the group factors; LOAD.GF = loadings on the general factors.

Table 4. Mean absolute error (MAE) across each variable level for each factor retention method.

Variable	Group factors							General factors			
	Kaiser		PA_{PAF}		PA_{PCA}		EGA_{LV}	PA_{PCA-FS}		EGA_{LV}	EGA_{LV-FS}
	K1	EKC	mean	$95th$	mean	$95th$		mean	$95th$		
MF											
zero	1.57	1.88	0.02	0.02	0.48	0.57	0.31	0.01	0.02	2.83	0.00
close	2.59	1.87	3.84	3.19	0.50	0.59	0.31	0.01	0.02	2.89	0.00
N											
500	4.87	4.04	0.44	0.27	1.14	1.37	0.31	0.04	0.07	3.42	0.00
1000	2.27	2.11	1.00	0.68	0.47	0.56	0.32	0.01	0.01	2.94	0.00
2000	0.86	0.98	2.12	1.72	0.23	0.26	0.31	0.00	0.00	2.64	0.00
5000	0.33	0.37	4.17	3.76	0.12	0.13	0.30	0.00	0.00	2.42	0.01
N.GF											
1	0.42	0.34	2.43	1.99	0.07	0.10	0.02	0.00	0.00	1.07	0.00
2	1.52	1.41	1.73	1.42	0.34	0.41	0.15	0.01	0.01	2.78	0.00
3	3.48	3.11	1.88	1.60	0.85	0.99	0.61	0.02	0.04	3.82	0.01
COR.GF											
0	1.62	1.61	1.24	0.99	0.40	0.48	0.28	0.01	0.02	2.49	0.00
.30	2.77	2.27	2.97	2.54	0.62	0.73	0.35	0.02	0.03	3.40	0.00
VAR.GRF											
4	0.15	4.32	0.91	0.67	1.46	1.73	0.77	0.04	0.08	2.17	0.01
6	0.99	1.84	1.66	1.34	0.30	0.37	0.29	0.00	0.00	3.04	0.00
8	2.55	0.88	2.31	1.94	0.11	0.13	0.11	0.00	0.00	3.15	0.00
10	4.64	0.47	2.85	2.47	0.09	0.09	0.05	0.00	0.00	3.07	0.00
NUM.GRF											
4	1.24	1.14	1.60	1.29	0.26	0.33	0.18	0.01	0.02	2.27	0.01
5	2.03	1.83	1.93	1.60	0.47	0.56	0.31	0.01	0.02	2.79	0.00
6	2.98	2.66	2.27	1.93	0.73	0.86	0.44	0.01	0.02	3.51	0.00
CROSS.GRF											
0	2.06	1.74	2.00	1.67	0.38	0.47	0.01	0.01	0.02	2.88	0.00
.15	2.06	1.90	1.96	1.64	0.49	0.59	0.15	0.01	0.03	2.93	0.00
.30	2.12	1.99	1.83	1.51	0.60	0.69	0.76	0.01	0.02	2.76	0.01
LOAD.GRF											
low	2.98	2.79	1.93	1.57	0.85	1.01	0.50	0.02	0.04	2.70	0.01
medium	1.19	0.96	1.94	1.64	0.13	0.15	0.11	0.00	0.00	3.01	0.00
LOAD.GF											
low	2.97	1.59	1.69	1.35	0.28	0.35	0.29	0.00	0.00	3.07	0.00
medium	1.20	2.16	2.18	1.86	0.70	0.81	0.32	0.02	0.04	2.64	0.01
Total	2.08	1.88	1.93	1.61	0.49	0.58	0.31	0.01	0.02	2.86	0.00

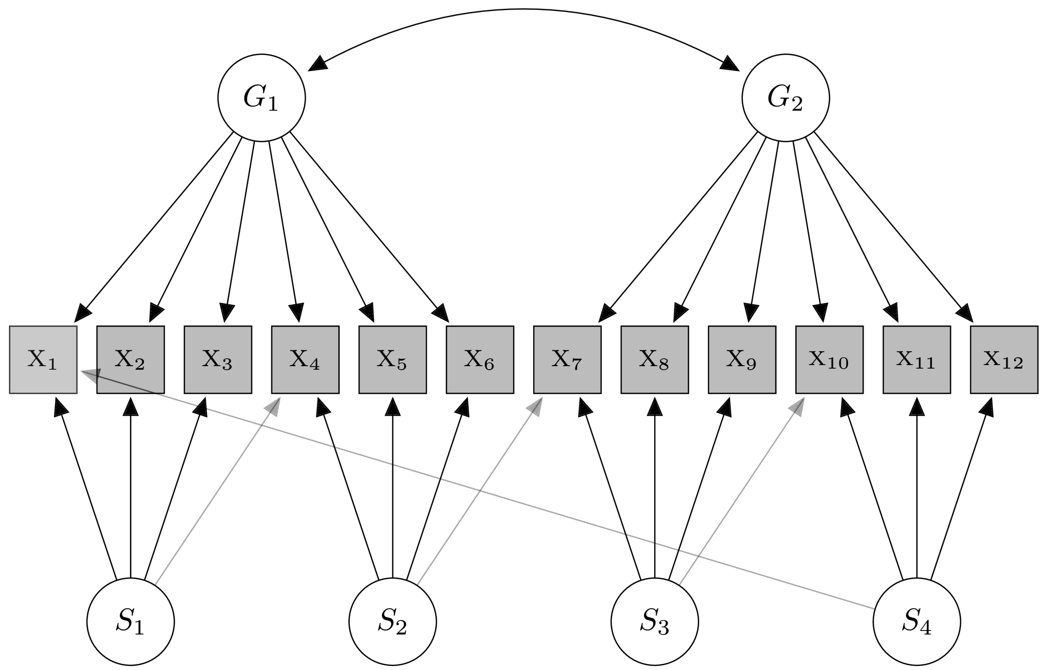
Note. K1 = Kaiser eigenvalue greater-than-one criterion; EKC = Empirical Kaiser Criterion; PA_{PAF} = Parallel analysis with principal axis factoring; PA_{PCA} = Parallel analysis with principal components; PA_{PCA-FS} = Parallel analysis with principal components on the first-order factor scores; EGA = Exploratory Graph Analysis; EGA_{LV} = EGA with Louvain; EGA_{LV-FS} = EGA with Louvain on the first-order factor scores; MF = population misfit; N = sample size; N.GF = number of general factors; COR.GF = correlation between general factors; VAR.GRF = number of indicators per group factor; NUM.GRF = number of group factors per general factor; CROSS.GRF = cross-loadings in the group factors; LOAD.GRF = loadings on the group factors; LOAD.GF = loadings on the general factors.

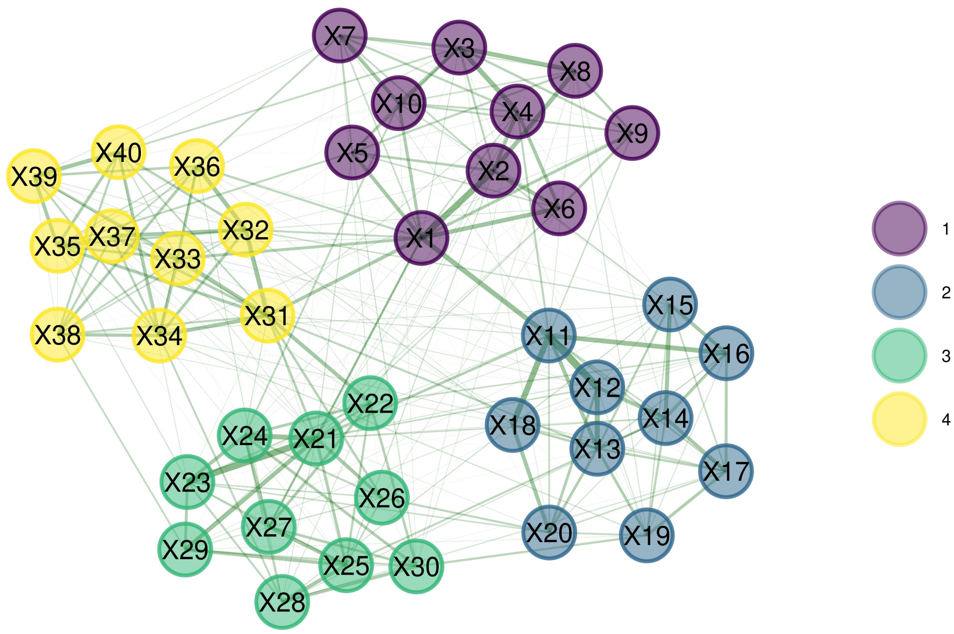
Table 5. Partial omega squared coefficients (Ω^2) from the ANOVAs on the absolute error for all the nine main effects, and for the remaining coefficients whose $\Omega^2 \geq .14$ or close in at least one factor retention method.

Coefficients	Group factors		General factors	
	PA$_{\text{PCA}}$	EGA$_{\text{LV}}$	PA$_{\text{PCA-FS}}$	EGA$_{\text{LV-FS}}$
Main effects				
VAR.GRF	**0.57**	**0.22**	0.03	0.01
N	**0.39**	0.00	0.02	0.00
N.GF	**0.29**	**0.18**	0.01	0.00
LOAD.GF	**0.15**	0.00	0.01	0.00
LOAD.GRF	**0.35**	0.12	0.01	0.00
NUM.GRF	**0.13**	0.04	0.00	0.00
MF	0.00	0.00	0.00	0.00
COR.GF	0.00	0.00	0.00	0.00
CROSS.GRF	0.03	**0.28**	0.00	0.01
Two-way interactions				
VAR.GRF × LOAD.GRF	**0.45**	0.09	0.03	0.01
VAR.GRF × N	**0.44**	0.01	0.06	0.00
VAR.GRF × N.GF	**0.28**	**0.14**	0.02	0.01
N × LOAD.GRF	**0.26**	0.00	0.02	0.00
N × N.GF	**0.20**	0.00	0.01	0.00
VAR.GRF × LOAD.GF	**0.19**	0.00	0.02	0.00
N.GF × LOAD.GRF	**0.16**	0.08	0.01	0.00
VAR.GRF × CROSS.GRF	0.05	**0.33**	0.00	0.02
N.GF × CROSS.GRF	0.01	**0.20**	0.00	0.01
Three-way interactions				
VAR.GRF × N × LOAD.GRF	**0.21**	0.00	0.06	0.00
VAR.GRF × N × N.GF	**0.18**	0.01	0.04	0.00
VAR.GRF × N.GF × LOAD.GRF	**0.17**	0.05	0.02	0.01
VAR.GRF × N.GF × CROSS.GRF	0.01	**0.22**	0.00	0.02
VAR.GRF × LOAD.GRF × CROSS.GRF	0.01	**0.13**	0.00	0.01

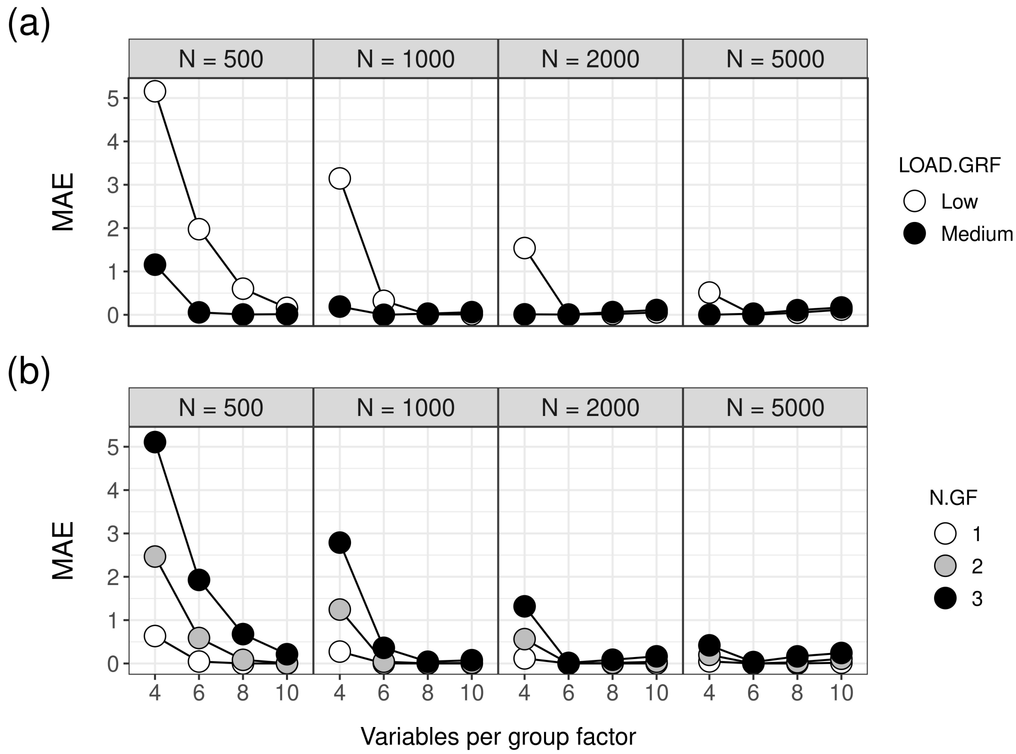
Note. PA$_{\text{PCA}}$ = Parallel analysis with principal components; PA$_{\text{PCA-FS}}$ = Parallel analysis with principal components on the first-order factor scores; EGA = Exploratory graph analysis; EGA$_{\text{LV}}$ = EGA with Louvain; EGA$_{\text{LV-FS}}$ = EGA with Louvain on the first-order factor scores. MF = population misfit; N.GF = number of general factors; COR.GF = Correlation between general factors; N = sample size; VAR.GRF = Number of indicators per group factor; NUM.GRF = number of group factors per general factor; CROSS.GRF = cross-loadings in the group factors; LOAD.GF = loadings on the general factors; LOAD.GRF = loadings on the group factors;

Figures

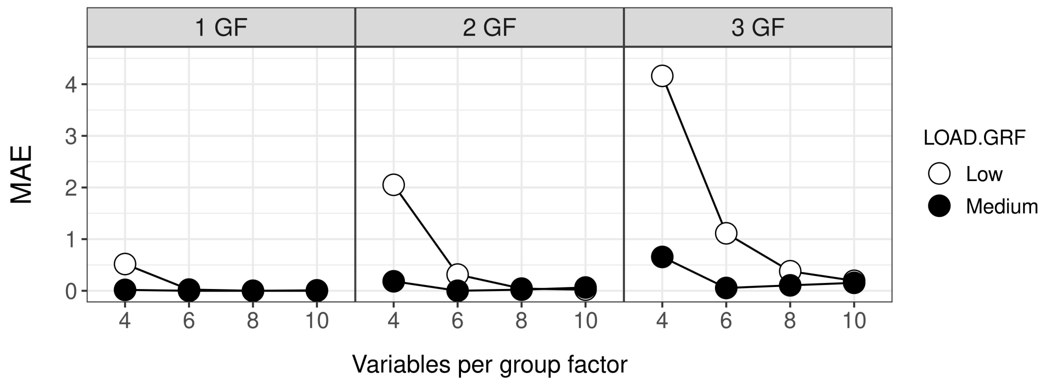




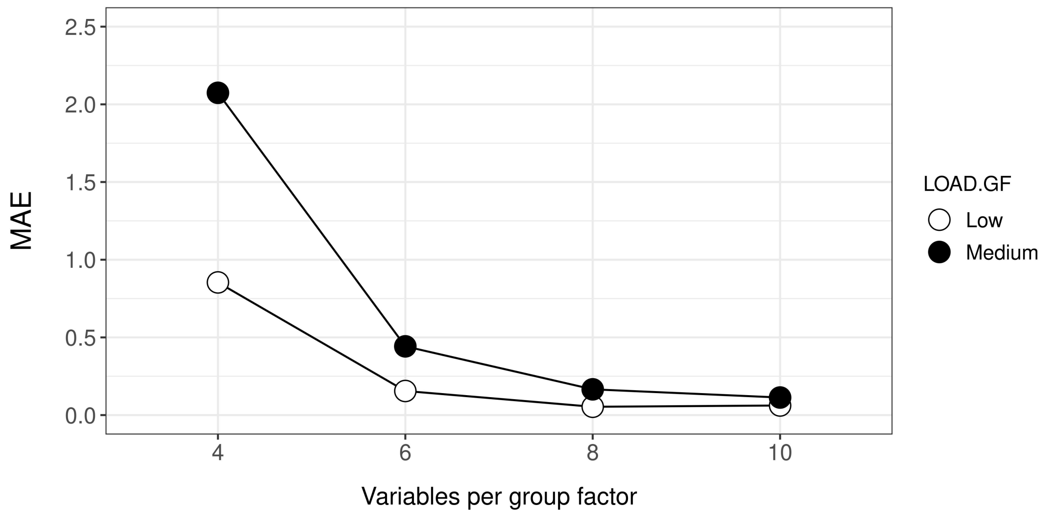
X7
X3
X8
X10
X4
X9
X5
X2
X6
X1
X40
X36
X39
X37
X32
X35
X33
X31
X38
X34
X15
X11
X16
X12
X22
X14
X18
X24
X21
X13
X17
X23
X26
X27
X20
X19
X29
X25
X30
X28
1
2
3
4



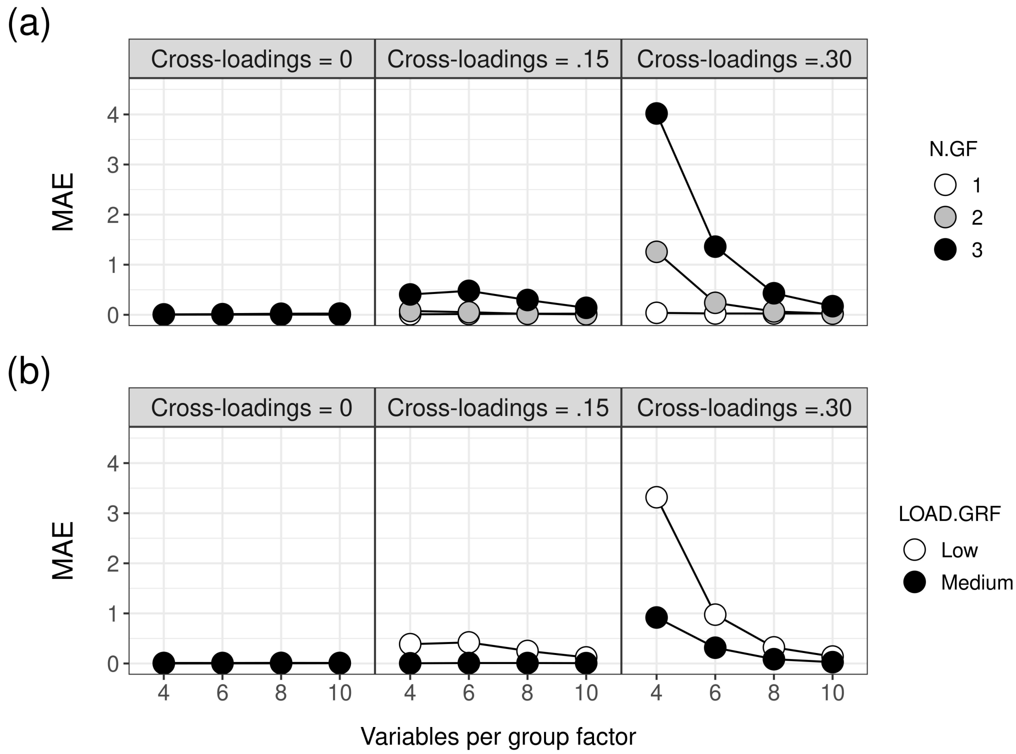
(a)
N = 500
N = 1000
N = 2000
N = 5000
MAE
LOAD.GRF
Low
Medium
(b)
N = 500
N = 1000
N = 2000
N = 5000
MAE
N.GF
1
2
3
Variables per group factor



1 GF
2 GF
3 GF
MAE
0
1
2
3
4
4
6
8
10
Variables per group factor
LOAD.GRF
Low
Medium



MAE
2.5
2.0
1.5
1.0
0.5
0.0
4
6
8
10
Variables per group factor
LOAD.GF
Low
Medium



(a)
Cross-loadings = 0
Cross-loadings = .15
Cross-loadings =.30
MAE
N.GF
1
2
3
(b)
Cross-loadings = 0
Cross-loadings = .15
Cross-loadings =.30
MAE
LOAD.GRF
Low
Medium
Variables per group factor

Figure captions

Figure 1. Illustration of a generalized bi-factor model with two general factors (G) and four group factors (S) for twelve indicators (X). The grey arrows represent cross-loadings among the group factors, with each group factor having an indicator that cross-load on another group factor.

Figure 2. Network estimated via a Gaussian Graphical Model with GLASSO. Each color represents a factor. Items were clustered with the Louvain algorithm.

Figure 3. Three-way interactions VAR.GRF $\times$ N $\times$ LOAD.GRF ($\Omega^2 = 0.22$) and VAR.GRF $\times$ N $\times$ N.GF ($\Omega^2 = 0.18$) for PA_{PCA}.

Figure 4. Three-way interaction VAR.GRF $\times$ N.GF $\times$ LOAD.GRF ($\Omega^2 = 0.16$) for PA_{PCA}.

Figure 5. Two-way interaction VAR.GRF $\times$ LOAD.GF ($\Omega^2 = 0.19$) for PA_{PCA}.

Figure 6. Three-way interactions VAR.GRF $\times$ CROSS.GRF $\times$ N.GF ($\Omega^2 = 0.22$) and VAR.GRF $\times$ CROSS.GRF $\times$ LOAD.GRF ($\Omega^2 = 0.13$) for EGA_{LV}.

Appendix

Table A1. Marginal fit indices for each variable level. The mean value is displayed in bold, and the single worst fit value is displayed in parentheses.

Variable	SRMR	RMSEA	CFI	Absolute residuals
N.GF				
1	**.0209** (.0263)	**.0266** (.0298)	**.9933** (.9902)	**.0711** (.0998)
2	**.0209** (.0266)	**.0217** (.0288)	**.9869** (.9801)	**.0803** (.0997)
3	**.0209** (.0263)	**.0214** (.0274)	**.9808** (.9702)	**.0789** (.0999)
COR.GF				
0	**.0209** (.0266)	**.0219** (.0298)	**.9865** (.9702)	**.0777** (.0999)
.30	**.0209** (.0264)	**.0216** (.0281)	**.9846** (.9732)	**.0782** (.0995)
VAR.GRF				
4	**.0209** (.0266)	**.0224** (.0298)	**.9851** (.9702)	**.0739** (.0988)
6	**.0209** (.0264)	**.0218** (.0288)	**.9857** (.9716)	**.0774** (.0999)
8	**.0209** (.0261)	**.0215** (.0273)	**.9860** (.9730)	**.0793** (.0993)
10	**.0209** (.0259)	**.0214** (.0270)	**.9862** (.9727)	**.0809** (.0998)
NUM.GRF				
4	**.0209** (.0266)	**.0220** (.0298)	**.9868** (.9754)	**.0758** (.0997)
5	**.0209** (.0263)	**.0217** (.0291)	**.9857** (.9729)	**.0783** (.0994)
6	**.0209** (.0263)	**.0216** (.0293)	**.9847** (.9702)	**.0795** (.0999)
CROSS.GRF				
0	**.0209** (.0262)	**.0217** (.0298)	**.9857** (.9702)	**.0771** (.0999)
.15	**.0209** (.0264)	**.0218** (.0294)	**.9859** (.9717)	**.0778** (.0995)
.30	**.0209** (.0266)	**.0218** (.0295)	**.9856** (.9713)	**.0787** (.0998)
LOAD.GRF				
low	**.0187** (.0224)	**.0194** (.0248)	**.9874** (.9770)	**.0707** (.0988)
medium	**.0231** (.0266)	**.0241** (.0298)	**.9841** (.9702)	**.0851** (.0999)
LOAD.GF				
low	**.0186** (.0220)	**.0194** (.0252)	**.9840** (.9702)	**.0704** (.0992)
medium	**.0231** (.0266)	**.0241** (.0298)	**.9875** (.9780)	**.0854** (.0999)
Total	**.0209** (.0266)	**.0218** (.0298)	**.9857** (.9702)	**.0779** (.0999)

Note. N.GF = number of general factors; COR.GF = correlation between general factors; VAR.GRF = number of indicators per group factor; NUM.GRF = number of group factors per general factor; CROSS.GRF = cross-loadings in the group factors; LOAD.GRF = loadings on the group factors; LOAD.GF = loadings on the general factors.

Table A2. Partial omega squared coefficients (Ω^2) from the ANOVAs on the absolute error for the recovery of the group factors for all the nine main effects, and for the remaining coefficients whose $\Omega^2 \geq .14$ in at least one factor retention method.

Variable	Kaiser		PA_{PAF}		PA_{PCA}		EGA_{LV}
	K1	EKC	mean	95*th*	mean	95*th*	
Main effects							
VAR.GRF	**0.84**	**0.84**	**0.30**	**0.30**	**0.57**	**0.62**	**0.22**
N	**0.84**	**0.82**	**0.62**	**0.64**	**0.39**	**0.46**	0.00
N.GF	**0.72**	**0.74**	0.05	0.04	**0.29**	**0.31**	**0.18**
LOAD.GF	**0.58**	**0.17**	0.05	0.06	**0.15**	**0.16**	0.00
LOAD.GRF	**0.58**	**0.67**	0.00	0.00	**0.35**	**0.40**	0.12
NUM.GRF	**0.47**	**0.48**	0.06	0.06	**0.13**	**0.15**	0.04
MF	**0.31**	0.00	**0.75**	**0.71**	0.00	0.00	0.00
COR.GF	0.09	0.00	**0.47**	**0.45**	0.00	0.00	0.00
CROSS.GRF	0.00	0.02	0.00	0.00	0.03	0.03	**0.28**
Two-way interactions							
VAR.GRF × N	**0.75**	**0.53**	**0.36**	**0.36**	**0.44**	**0.48**	0.01
N × N.GF	**0.63**	**0.58**	0.07	0.05	**0.20**	**0.23**	0.00
VAR.GRF × N.GF	**0.60**	**0.52**	0.06	0.05	**0.28**	**0.30**	**0.14**
N × LOAD.GF	**0.44**	0.02	0.03	0.03	0.06	0.07	0.00
VAR.GRF × LOAD.GRF	**0.42**	**0.35**	0.00	0.00	**0.45**	**0.48**	0.09
VAR.GRF × LOAD.GF	**0.42**	0.08	0.00	0.00	**0.19**	**0.19**	0.00
N × LOAD.GRF	**0.41**	**0.34**	0.00	0.00	**0.26**	**0.30**	0.00
VAR.GRF × NUM.GRF	**0.34**	**0.22**	0.00	0.00	0.11	0.12	0.02
N × NUM.GRF	**0.32**	**0.30**	0.03	0.04	0.9	0.11	0.00
N.GF × LOAD.GRF	**0.30**	**0.32**	0.04	0.03	**0.16**	**0.17**	0.08
N.GF × LOAD.GF	**0.30**	0.04	0.02	0.01	0.05	0.05	0.00
VAR.GRF × MF	**0.20**	0.00	**0.31**	**0.32**	0.01	0.01	0.00
N.GF × NUM.GRF	**0.18**	**0.23**	0.02	0.01	0.06	0.06	0.02
LOAD.GF × LOAD.GRF	**0.14**	0.03	0.00	0.00	0.09	0.09	0.00
N × MF	0.02	0.00	**0.62**	**0.64**	0.00	0.00	0.00
MF × COR.GF	0.10	0.00	**0.47**	**0.45**	0.00	0.00	0.00
N × COR.GF	0.00	0.00	**0.36**	**0.39**	0.00	0.00	0.00
VAR.GRF × COR.GF	0.04	0.00	**0.13**	**0.13**	0.00	0.00	0.00
VAR.GRF × CROSS.GRF	0.00	0.01	0.00	0.00	0.05	0.05	**0.33**
N.GF × CROSS.GRF	0.00	0.00	0.00	0.00	0.01	0.00	**0.20**
Three-way interactions							
VAR.GRF × N × N.GF	**0.47**	**0.18**	0.03	0.02	**0.18**	**0.19**	0.01
VAR.GRF × N × LOAD.GF	**0.22**	0.03	0.00	0.00	0.04	0.04	0.00
VAR.GRF × N × LOAD.GRF	**0.20**	**0.31**	0.00	0.00	**0.21**	**0.23**	0.00
VAR.GRF × N × NUM.GRF	**0.18**	0.05	0.01	0.01	0.07	0.07	0.00
VAR.GRF × N.GF × LOAD.GRF	**0.16**	0.04	0.01	0.01	**0.17**	**0.16**	0.05
N × N.GF × LOAD.GF	**0.16**	0.00	0.00	0.00	0.02	0.02	0.00
VAR.GRF × N.GF × LOAD.GF	**0.15**	0.01	0.00	0.00	0.04	0.04	0.00
N × N.GF × LOAD.GRF	**0.13**	0.09	0.00	0.00	0.11	0.11	0.00
N × MF × COR.GF	0.00	0.00	**0.36**	**0.39**	0.00	0.00	0.00
VAR.GRF × N × MF	0.02	0.00	**0.34**	**0.34**	0.00	0.00	0.00
VAR.GRF × N × COR.GF	0.00	0.00	**0.19**	**0.20**	0.00	0.00	0.00
VAR.GRF × MF × COR.GF	0.04	0.00	**0.13**	**0.13**	0.00	0.00	0.00
VAR.GRF × CROSS.GRF × N.GF	0.00	0.00	0.00	0.00	0.01	0.01	**0.22**
VAR.GRF × CROSS.GRF × LOAD.GRF	0.00	0.01	0.00	0.00	0.01	0.01	**0.13**

Note. K1 = Kaiser eigenvalue greater-than-one criterion; EKC = Empirical Kaiser Criterion; PA_{PAF} = Parallel analysis with principal axis factoring; PA_{PCA} = Parallel analysis with principal components; EGA_{LV} = Exploratory graph analysis with Louvain; MF = population misfit; N.GF = number of general factors; COR.GF = Correlation between general factors; N = sample size; VAR.GRF = Number of indicators per group factor; NUM.GRF = number of group factors per general factor; CROSS.GRF = cross-loadings in the group factors; LOAD.GF = loadings on the general factors; LOAD.GRF = loadings on the group factors.