Dimensionality Assessment in the Presence of Wording Effects: A Network Psychometric and Factorial Approach

Alejandro Garcia-Pardina[1], Francisco J. Abad[1], Alexander P. Christensen[2], Hudson Golino[3], and Luis Eduardo Garrido[4]

[1]Universidad Autónoma de Madrid, Spain

[2]Vanderbilt University, USA

[3]University of Virginia, USA

[4]Pontificia Universidad Católica Madre y Maestra, Dominican Republic

Author Note

Francisco J. Abad https://orcid.org/0000-0001-6728-2709

Alexander P. Christensen https://orcid.org/0000-0002-9798-7037

Hudson Golino https://orcid.org/0000-0002-1601-1447

Luis Eduardo Garrido https://orcid.org/0000-0001-8932-6063

We have no conflicts of interest to disclose.

This study was not preregistered. All R simulation codes, results, and the empirical data are available at https://osf.io/f92hu/?view_only=304cc8a3c0434f95914badde7a4fbfea

Correspondence concerning this article should be addressed to Dr. Luis Eduardo Garrido, Vicerrectoría de Investigación e Innovación, Edificio Padre José Luis Alemán, Pontificia Universidad Católica Madre y Maestra (PUCMM), Av. Simón Bolívar, 10109, Santo Domingo, Dominican Republic. email: luisgarrido@pucmm.edu.do

Abstract

This study proposes a procedure for substantive dimensionality estimation in the presence of wording effects, the inconsistent response to regular and reversed self-report items. The procedure developed consists of subtracting an approximate estimate of the wording effects variance from the sample correlation matrix and then estimating the substantive dimensionality on the residual correlation matrix. This is achieved by estimating a random intercept factor with unit loadings for all the regular and unrecoded reversed items. The accuracy of the procedure was evaluated through a broad simulation study that manipulated six relevant variables and employed the exploratory graph analysis (EGA) and parallel analysis (PA) retention methods. The results indicated that combining the proposed procedure with EGA or PA achieved high accuracy in estimating the substantive latent dimensionality, but that EGA was superior. Additionally, the present findings shed light on the complex ways that wording effects impact the dimensionality estimates when the response bias in the data is ignored. A tutorial on substantive dimensionality estimation with the R package *EGAnet* is offered, as well as practical guidelines for applied researchers.

Keywords: number of factors, dimensionality, wording effects, method factor, response bias, exploratory graph analysis, parallel analysis

Dimensionality Assessment in the Presence of Wording Effects: A Network Psychometric and Factorial Approach

Self-reports are one of the most prevalent means of collecting information in the behavioral and health sciences (Chan, 2010). The use of self-reports is advantageous because it enables the collection of large amounts of quantitative data at a low cost, which can then be used to develop generalizable findings that are highly useful for society (Demetriou et al., 2015). Despite their widespread use, however, the validity of self-reported data is threatened by response biases which can result from many causes associated with the rater, item characteristics, item context, and measurement context (Podsakoff et al., 2003). Of these response biases, those related to the semantic polarity of the items, usually referred to as "wording effects", are particularly prevalent and adverse (Swain et al., 2008; Weijters et al., 2013). Among their various negative impacts on validity, they can deteriorate model fit (Schmalbach et al., 2020), increase the dimensionality of the observed scores through the emergence of artifactual "method" factors (DiStefano & Motl, 2006), distort the factor loading structures (Garrido et al., 2022), alter structural relationships (Nieto et al., 2021), and reduce scale reliabilities (Vigil-Colet et al., 2020).

Wording effects are logically inconsistent answers to regular and reversed items of the same dimension (Kam, 2018). An item is regular-keyed if its semantic polarity is in the direction of the construct being measured (e.g., the item "Am afraid of many things" for Neuroticism). Conversely, an item is reverse-keyed if its semantic polarity is in the opposite direction of the construct (e.g., the item "Remain calm under pressure" for Neuroticism). Respondents that agree (or disagree) with both regular and reversed-keyed items of the same dimension are thought to be answering in a logically inconsistent manner (Nieto et al., 2021). Research suggests that these wording effects result from response biases, such as carelessness, acquiescence, and comprehension difficulty (Baumgartner et al., 2018; Swain et al., 2008; Weijters et al., 2013). *Carelessness* occurs when respondents pay insufficient attention to the items' content, producing random or nonrandom response patterns that are unrelated to the measures being administered. The *acquiescence* response bias is the tendency to express agreement with statements regardless of their content. *Comprehension difficulty* arises when the person's response does not match their true beliefs because of problems in properly understanding the items' content or in selecting the appropriate response options.

Determining the number of factors to retain is one of the most important decisions in factor analytical research (Hayton et al., 2004; Henson et al., 2006). Specifying less –underfactoring– or more –overfactoring– dimensions than those present in the population can have detrimental effects on the quality of the factor solutions, including substantial error in the factor loading and factor score estimates, factor splitting, inadmissible solutions, and the emergence of uninterpretable factors (Auerswald & Moshagen, 2019). The inherent difficulties of dimensionality assessment are further exacerbated by the response biases related to wording effects, which tend to increase the latent dimensionality of observed scores (García-Batista et al., 2021; Kam, 2018; Schmalbach et al., 2020; Yang et al., 2018; Zhang & Savalei, 2016). This, in turn, can lead to long-standing controversies regarding the substantive or artifactual nature of dimensions underlying mixed-worded scales (Gnambs et al., 2018).

At present, dimensionality assessment methods are limited because they only inform of the *total* dimensionality underlying the observed data, offering no guidance regarding whether one or more of the suggested factors are due to wording effects. This can make researchers over-rely on theory and model comparisons that may not result in the selection of the optimal models, particularly when multiple substantial factors and multiple wording factors underlie the data

(Kam, 2018). This situation is especially adverse in more exploratory contexts where theory regarding the latent structure is more tenuous. Furthermore, for dimensionality methods that also provide an estimate of the latent structure (i.e., item groupings), such as exploratory graph analysis (EGA; Golino & Epskamp, 2017), wording effects can lead to non-interpretable solutions (Juárez-García et al., 2021).

In this paper we propose a procedure for *substantive dimensionality estimation* to address the problems related to dimensionality assessment in the presence of wording effects. The accuracy of the proposed procedure is tested via a broad Monte Carlo simulation study and an empirical study with Big Five personality data. Additionally, we conduct the first in-depth systematic evaluation of the performance of traditional dimensionality methods with simulated data contaminated by wording effects. To contextualize the study, the rest of the Introduction is organized as follows: first, we present a brief overview of two highly recommended dimensionality assessment methods, which are the focus of this study. Second, we provide a detailed description and rationale for the proposed procedure. Third, we summarize the aims and hypotheses of the study.

Dimensionality Assessment Methods

Many methods have been proposed to assess the latent dimensionality of psychological data (for a review, see Auerswald & Moshagen, 2019, and Goretzko et al., 2021). Of these, parallel analysis (PA; Horn, 1965) and EGA (Golino & Epskamp, 2017) have shown excellent performance across numerous studies (e.g., Auerswald & Moshagen, 2019; Cosemans et al., 2021; Garrido et al., 2013, 2016; Golino et al., 2020; Golino & Demetriou, 2017; Xia, 2021). As a result, PA and EGA are currently some of the most recommended dimensionality assessment methods (Cosemans et al., 2021; Ferrando et al., 2022; Goretzko et al., 2021).

Parallel Analysis

Horn (1965) proposed PA as a sample alternative to the eigenvalue-greater-than-one rule, which posits that factors with eigenvalues > 1 should be retained (Kaiser, 1960). Horn argued that the rank of a sample correlation matrix should be estimated by subtracting out the component in the eigenvalues due to sampling error and capitalization on chance (Horn, 1965). To accomplish this, Horn suggested generating random data from a null model of zero correlations and with the same number of variables and sample size as the data under assessment. According to the PA procedure, factors should then be retained if their eigenvalues are greater than the eigenvalues of the corresponding factors from this generated data. It follows, therefore, that at the population level PA and the eigenvalue-greater-than-one rule are equivalent, as the eigenvalues from a null correlation matrix are all equal to one.

Several variations of the PA method have been proposed (for a review, see Lim & Jahng, 2019). These variants suggest using different extraction methods, generating the random data from different models, or employing different criteria to aggregate the random eigenvalues. However, there is converging evidence that Horn's original formulation of computing the eigenvalues from the full correlation matrix (i.e., principal component analysis [PCA] eigenvalues) and generating the data from a null model might provide the best performing formulation (Garrido et al., 2013; Lim & Jahng, 2019; Xia, 2021). In terms of the aggregating criteria for the random eigenvalues, the mean tends to perform better for correlated structures, while the 95th percentile is superior for orthogonal structures or data with population error (Garrido et al., 2013; Lim & Jahng, 2019; Xia, 2021). Additionally, it is important that the random data preserves the distributional properties of the assessed data to ensure the effectiveness of all PA variants (Lubbe, 2019). This can be achieved through various procedures, such as random column permutations of the original data or discretization of multivariate normal data using the item thresholds (Garrido et al., 2013; Lubbe, 2019).

Exploratory Graph Analysis

EGA (Golino & Epskamp, 2017) is a network psychometrics approach to identify communities comparable to factors in a network. Psychometric networks represent variables as nodes (circles) and the relationships between nodes as edges (lines). The EGA procedure applies a network estimation method followed by a community detection algorithm for weighted networks (Golino et al., 2020). In the network psychometrics literature, the most common method to estimate a network is to use the graphical least absolute shrinkage and selection operator (GLASSO; Friedman et al., 2014). The GLASSO seeks to maximize the penalized ℓ_1-norm log-likelihood by shrinking parameter estimates of the inverse covariance matrix and setting some to zero. After estimation, the inverse covariance matrix is converted to a partial correlation matrix. The GLASSO has a rho parameter that is selected using the extended Bayesian information criterion (EBIC; Epskamp & Fried, 2018).

After the network is estimated, EGA applies a community detection algorithm to identify communities or sets of nodes that are more connected with themselves than other nodes in the network. The original algorithm applied in the EGA approach is the Walktrap algorithm, which implements random walks from node to node in the network and identifies communities based on how walks tend to stick within sets of nodes (Pons & Latapy, 2005). More recently, EGA has been evaluated using other community detection algorithms, finding that the Louvain algorithm (Blondel et al., 2008) had the highest accuracy for recovering the population number of factors in data generated from factor models (Christensen et al., 2021). The Louvain algorithm identifies communities hierarchically, starting with smaller clusters or sets of nodes and merging them into larger and larger clusters by optimizing a criterion called modularity (the extent to which nodes are more connected to nodes within their community than to nodes in other communities; Gates et al., 2016). This feature allows the Louvain algorithm to output community solutions at different levels (i.e., smaller clusters and larger clusters) analogous to lower-order and higher-order structures often seen in hierarchical factor models. The Walktrap algorithm also identifies communities hierarchically using Ward's method and decides on the best composition of nodes into communities by choosing a solution that maximizes modularity (Pons & Latapy, 2005). However, the final solution of the Walktrap algorithm is solely the partition of nodes into communities in which modularity is the highest. On the other side, the Louvain algorithm uses a very different approach, with two phases: one where modularity is optimized by allowing only local changes of communities and one where these estimated communities are aggregated to build a new network of communities. The process is repeated iteratively until no increase in modularity is possible (Blondel et al., 2008).

A Proposal for Substantive Dimensionality Estimation

In this section we propose a procedure to estimate the number of substantive factors underlying mixed-worded scales. The procedure consists of first subtracting an approximate estimate of the wording effects variance from the sample correlation matrix and then computing the dimensionality estimates on the obtained residual correlation matrix. For this, we leaned on the random intercept item factor analysis (RIIFA) model developed by Maydeu-Olivares and Coffman (2006), which has shown excellent performance in accounting for the wording effects' variance across simulation (e.g., de la Fuente & Abad, 2020; Garrido et al., 2022; Savalei & Falk, 2014; Nieto et al., 2021) and empirical (e.g., Aichholzer, 2014; Arias et al., 2020; Schmalbach et al., 2020; Weydmann et al., 2020) studies. Notably, the RIIFA model has been produced excellent levels of fit for data contaminated with different types of wording effects (Nieto et al., 2021), as well as a good recovery of the substantive factor loadings, even with unequal amounts of wording effects across items (Savalei & Falk, 2014). We will first introduce the RIIFA model, and then we will describe why it is well suited to address the problem of dimensionality assessment in the presence of wording effects.

For an m-dimensional common factor model, the response of participant j to item i, y_{ij}, can be written as

$$y_{ij} = \mu_i + \boldsymbol{\lambda}_i' \boldsymbol{\eta}_j + e_{ij}, \tag{1}$$

where μ_i is the intercept for item i, $\boldsymbol{\lambda}_i$ is the vector of factor loadings for item i, $\boldsymbol{\eta}_j$ is a vector of factor scores for participant j, and e_{ij} is the error term for participant j on item i. As can be seen in Equation 1, the intercept μ_i does not change across participants.

Employing the typical assumption of the common factor model that the error terms are uncorrelated with the common factors, the model-implied covariance matrix, Σ_y, can be expressed as

$$\Sigma_y = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{2}$$

where $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of factor loadings, $\boldsymbol{\Psi}$ is the $m \times m$ factor covariance matrix, $\boldsymbol{\Theta}$ is the $p \times p$ error covariance matrix, and p is the number of items.

In the common factor model with a random intercept factor, the assumption of intercepts common to all respondents is relaxed by allowing the intercepts to vary from respondent to respondent (Maydeu-Olivares & Coffman, 2006). Thus, In the RIIFA model the intercept γ_{ij} is partitioned into a fixed part, μ_i, that is common across respondents but varies across items, and a random part, ζ_j, that varies across respondents but is common to all items. The term ζ_j allows the RIIFA model to account for participants' idiosyncratic use of the response scale that is common across items. The equation for y_{ij} in the RIIFA model can thus be written as

$$y_{ij} = \gamma_{ij} + \boldsymbol{\lambda}_i'\boldsymbol{\eta}_j + e_{ij}, \qquad \gamma_{ij} = \mu_i + \zeta_j. \tag{3}$$

If, in addition to the typical assumptions of the common factor model, we assume that the mean of ζ is zero, that ζ is uncorrelated with e, and that ζ is uncorrelated with the common factors, then Σ_y can be expressed as

$$\Sigma_y = \mathbf{1}\varphi\mathbf{1}' + \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{4}$$

where $\mathbf{1}$ is a $p \times 1$ vector of ones and φ is the variance of the random component of the intercept, ζ.

As can be seen in Equation 4, the RIIFA model does not estimate the intercepts for each respondent, instead, it estimates their variance, φ. The term $\mathbf{1}\varphi\mathbf{1}'$ is estimated by simply adding an additional factor to the m-dimensional factor model, where the unstandardized factor loadings of all the items are fixed to 1 (assuming that the reversed items are unrecoded), and φ is freely estimated. This additional factor is specified to be orthogonal to the m common factors.

For the current proposal, we are interested only in the additional factor that corresponds to the random intercept. Considering that with unrecoded reversed items the loadings of the regular and reversed items on the substantive factors would have opposite signs, having all the loadings be equal on the random intercept factor suggest an artifactual relationship (of opposite direction than suggested by substantive theory) between the two groups of items. It is this property of the random intercept factor, in addition to its orthogonality to the substantive factors, that could potentially allow for a useful approximate estimate of the wording effects variance independent of the rest of the model, which in the current context of dimensionality assessment is unknown (i.e., the number of substantive common factors). Based on this, we propose to estimate a residual $p \times p$ correlation matrix, $\mathbf{R_R}$, by subtracting the model-implied correlation matrix of the random intercept factor from the $p \times p$ sample correlation matrix, $\mathbf{R_S}$, such that

$$\mathbf{R_R} = \mathbf{R_S} - \mathbf{1}\varphi\mathbf{1}' + \mathbf{A}\,, \tag{5}$$

where $\mathbf{A}$ is a $p \times p$ diagonal matrix that ensures unit values for the diagonal elements in $\mathbf{R_R}$.

The next step is to estimate the number of substantive factors by applying the dimensionality procedures on $\mathbf{R_R}$. The EGA method can be directly computed on the residual correlation matrix, given that it is positive definite. Therefore, if $\mathbf{R_R}$ is not positive definite, it is first smoothed. In the case of PA, the empirical eigenvalues would be computed from $\mathrm{R_R}$, while the criterion eigenvalues would be computed in the usual way (e.g., from the correlation matrices of randomly permuted sample data).

An important consideration in the development of the procedure is that in some instances it is necessary to model more than one method factor to account for the wording effects in the data. Particularly, with data composed of multiple substantial factors it may be necessary to model as many as one random intercept factor for each mixed-worded substantive factor (Garrido et al., 2022; Kam, 2018). This presents a problem because in an exploratory context where dimensionality is being assessed, the number of substantive factors, and their composition, is part of the question being ascertained. Thus, we propose to model a single random intercept for all variables in a dataset, which would not require any prior knowledge or input by the researcher. This is based on the expectation that extracting a single random intercept factor might be enough to recover the substantive dimensionality in the data, even in cases where multiple wording factors are optimal. We expect that this single random intercept factor would provide a useful approximation to the wording variance, such that when subtracted from the correlation matrix, would weaken the method factors enough to no longer be detected by the retention methods.

Previous research suggests that if a random intercept factor is estimated on data weakly or uncontaminated of wording effects, the model either does not converge, or estimates very low loadings on the random intercept factor (Garrido et al., 2022). Therefore, we propose that nonconvergence of the random intercept model be interpreted as a signal of null or trivial wording variance and that the dimensionality estimates be computed in the usual way based on the sample correlation matrix. In those cases where the estimate of the random intercept factor does converge for uncontaminated data, we do not expect any meaningful loss in accuracy from our proposed method, as the amount of variance subtracted from the sample correlation matrix would generally be negligible. To summarize, Table A1 in the Appendix outlines the steps of the proposed procedure for substantive dimensionality estimation.

The Present Study

The current study was composed of a broad simulation study that aimed to (1) examine the impact of wording effects on the performance of traditional factor retention methods, and (2) examine the impact of wording effects on the performance of the same methods accompanied by our proposed procedure for substantive dimensionality estimation. In addition to the simulation, the methods were tested using an empirical dataset of Goldberg's Big Five markers. The retention methods evaluated were PA and EGA.

Based on the cited literature considering the impact of wording effects on factor models three hypotheses were postulated:

H1: the methods employing the residual correlations will provide approximately equal estimates to the methods employing the sample correlations in the absence wording effects.

H2: the methods employing the sample correlations will estimate more factors in the presence of wording effects, providing overestimates of the number of substantive factors.

H3: the methods employing the residual correlations will provide more accurate estimates of the number of substantive factors in the presence of wording effects than the methods employing the sample correlations.

Method

Simulation Design

The performance of the factor retention methods was ascertained using Monte Carlo methods. The data were simulated from a model containing three substantive factors and two wording method factors (Figure 1). Two of the substantive factors were balanced, containing the same number of regular and reversed items, while the other substantive factor was composed of regular items only. Additionally, a wording factor was postulated for each of the substantive factors combining regular and reversed items. Because this model contains multiple wording factors, and because one of the substantive factors contains only regular items that are uncontaminated by wording effects, this is a particularly challenging test for the proposed method, which estimates a single random intercept factor with equal loadings across all items.

The simulation design included two variables related to the wording effects and four variables known to affect dimensionality estimates. It can be summarized as follows:

1. Wording factor loadings: WFL = .00, .15, .20, .25, and .30.
2. Wording factor correlations: WFC = .00, .30, .50, .70, and 1.00.
3. Substantive factor loadings: FL = .50, .60, and .70.
4. Substantive factor correlations: FC = .00, .30, and .50.
5. Variables per factor: VF = 6 and 10.
6. Sample size: N = 300, 500, and 1000.

PLEASE INSERT FIGURE 1 ABOUT HERE

The *wording factor loading* levels included the condition without wording effects (WFL = .00) to serve as a baseline, as well as the small and large levels of .15 and .30, which were considered in Savalei and Falk (2014). Additionally, we simulated the middle values of .20 and .25 to get a more nuanced perspective of the impact of this variable. The *wording factor correlation* levels were chosen to cover the range from orthogonal wording factors (WFC = .00) to wording factors with perfect convergence (WFC = 1.00). The latter condition is equivalent to simulating a single wording factor for all the indicators of the two balanced factors. *Substantive factor loadings* of .50, .60, and .70 may be considered poor, good, and very good, respectively (Comrey & Lee, 1992). The *substantive factor correlations* include the orthogonal condition (.00), as well as medium (.30) and large (.50) correlations (Cohen, 1992). Regarding the *variables per factor*, 6 and 10 can be considered as medium and large numbers for modern questionnaires. Moreover, six is the minimum to ensure that both the regular and reversed items sufficiently identify the balanced factors on their own (three indicators is the minimum to identify a factor). As far as *sample size*, values of 300, 500, and 1000 may be considered as small, medium, and large, respectively (Li, 2016).

The design of the study was not fully crossed because when the WFL were .00 there were no WFC levels. Therefore, the design may be broken down into two parts: the conditions without wording effects (WFL = .00) composed of a 3 × 3 × 2 × 3 (FL × FC × VF × N) design, producing 54 conditions, and the conditions with wording effects (WFL > .00) composed of a 4 × 5 × 3 × 3 × 2 × 3 (WFL × WFC × FL × FC × VF × N) design, producing 1,080 conditions. In total there were 1,134 conditions and 100 replicates for each, resulting in 113,400 data matrices.

Data Generation

The sample data matrices were generated according to the following common factor model procedure: first, the reproduced population correlation matrix (with communalities in the diagonal) was computed as:

$$\mathbf{R_D} = \mathbf{\Lambda \phi \Lambda}^{\mathrm{T}} \quad (6)$$

where $\mathbf{R_D}$ is the reproduced population correlation matrix, $\mathbf{\Lambda}$ is the population factor loading matrix, and $\boldsymbol{\phi}$ is the population factor correlation matrix.

The population correlation matrix $\mathbf{R_P}$ was then obtained by inserting unities in the diagonal of $\mathbf{R_D}$, thereby raising the matrix to full rank. The next step was performing an eigen decomposition of $\mathbf{R_P}$, such that:

$$\mathbf{R_P} = \mathbf{VDV}^{\mathrm{T}}, \quad (7)$$

where $\mathbf{V}$ is a matrix of eigenvectors and $\mathbf{D}$ is a diagonal matrix of eigenvalues of $\mathbf{R_P}$. The scale matrix $\mathbf{S}$ was then computed:

$$\mathbf{S} = \mathbf{VD}^{1/2} \quad (8)$$

Finally, the sample matrix of continuous variables **X** was subsequently computed as:

$$\mathbf{X} = \mathbf{SZ} \quad (9)$$

where $\mathbf{Z}$ is a matrix of random standard normal deviates with rows equal to the sample size and columns equal to the number of variables.

Dimensionality Estimation

Considering the different variants of PA and EGA, a total of eight factor retention methods were evaluated. The retention methods that employed the *sample* correlation matrices included PA with PCA eigenvalues (PA) and exploratory graph analysis with GLASSO estimation (EGA). For PA, the mean (PA_m) and 95th percentile (PA_{95}) eigenvalue aggregating criteria were employed, while for EGA the Walktrap (EGA_{WT}) and Louvain (EGA_{LV}) clustering algorithms were examined. For the methods employing the *residual* correlation matrix after subtracting a random intercept (*ri*) factor, these were labelled correspondingly: *ri*PA_m, *ri*PA_{95}, *ri*EGA_{WT}, and *ri*EGA_{LV}. If the EGA_{LV} suggested a hierarchical solution, the highest (more general) level solution was used.

The retention methods using the residual correlation matrices after the subtraction of the random intercept model-implied correlation matrices were computed following the procedure outlined in Introduction and summarized in Table A1 in the Appendix. When at least one of the retention methods returned an error in the dimensionality estimate the replicate was discarded and a new sample data matrix was generated. This occurrence was exceedingly rare. The PA_m and PA_{95} methods produced dimensionality estimates for all replicates. EGA_{WT}, EGA_{LV}, *ri*PA_m, and *ri*PA_{95} returned an error on one occasion (0.001%), and *ri*EGA_{WT}, and *ri*EGA_{LV} failed to produce estimates for 13 replicates (0.01%).

Assessment Criteria

The factor retention methods were evaluated in their capacity to estimate the number of *substantive* factors in the population. Three complementary criteria were used to assess their performance: the hit rate (HR), the mean bias error (MBE), and the mean absolute error (MAE). The HR ranges from 0 to 1 and indicates the proportion of cases where the retention method correctly estimates the number of substantive factors. The MBE is the average of the differences between the estimates provided by the retention methods and the number of substantive factors in the population. An MBE of 0 indicates a lack of bias in the estimations, while positive and negative values indicate estimates of more and less factors, respectively, than the number of substantive factors present in the population. The MAE is the average of the absolute differences between the estimates provided by the retention methods and the number of substantive factors in the population. A MAE of 0 indicates perfect estimations, while higher values indicate greater absolute differences between the estimated and population number of substantive factors.

To ascertain the impact of the manipulated variables, and their interactions, on the performance of the retention methods we performed analyses of variance (ANOVAs) for each method separately. The MAE criterion was specified as the dependent variable in the ANOVAs. The partial eta squared statistic (η_p^2) was used to measure the size of the effects, with values of .01, .06, and .14, considered as small, medium, and large, respectively (Cohen, 1992).

The descriptive statistics and ANOVAs were computed using the IBM SPSS software version 25. The simulation study was performed in the R programming language (version 4.0.3). The sample data was generated by inputting the population correlation matrices in the R function *rmvnorm* contained in the *mvtnorm* package (version 1.1-1; Genz et al., 2020). The EGA estimates were obtained using the function *EGA* contained in the R package *EGAnet* (version 1.0.0; Golino & Christensen, 2021). The leading eigenvalue unidimensionality check was employed for EGA (Christensen et al., 2021). The PA estimates were obtained from R code developed by the authors. Following Garrido et al. (2013), reference eigenvalues were computed for each combination of number of variables and sample size considered in the study design. For each combination, 1000 sample data matrices of normal variates were generated from a null population correlation matrix. The random intercept factors were estimated using the R package *lavaan* (version 0.6-9; Rosseel, 2012). The nonpositive definite residual correlation matrices were smoothed with the Knol and Berger method using the function *corSmooth* of the *fungible* R package (version 1.99; Waller et al., 2021). This study was not preregistered. All R simulation codes and results, and the empirical data used in this study are available at https://osf.io/f92hu/?view_only=304cc8a3c0434f95914badde7a4fbfea.

Results

Conditions without Wording Effects

The evaluation of conditions without wording effects served two objectives: (1) to determine baseline values for the effectiveness of the retention methods, and (2) to assess the performance of the proposed "random intercept" methods when employed with data uncontaminated by wording effects. Regarding the former, the traditional formulations of the PA and EGA methods achieved approximately perfect accuracy in the estimation of the number of substantive factors (0.999 ≤ HR ≤ 1.000, -0.002 ≤ MBE ≤ 0.000, 0.000 ≤ MAE ≤ 0.002). Regarding the latter, the results showed that the proposed *ri*PA and *ri*EGA methods also achieved approximately perfect accuracy (0.999 ≤ HR ≤ 1.000, -0.002 ≤ MBE ≤ 0.000, 0.000 ≤ MAE ≤ 0.003). These results indicate that inadvertently using the proposed random intercept methods with uncontaminated data does not negatively impact the accuracy of the dimensionality estimates, thus supporting hypothesis H1.

Conditions with Wording Effects

The performance of the factor retention methods across the levels of the wording effects variables and in total are shown in Table 1, while the results across the rest of the manipulated variables are shown in Table 2. As can be seen in the tables, the mean and 95th percentile criteria for PA produced a very similar pattern of results, with the 95th percentile criterion being slightly superior. Likewise, the Walktrap and Louvain clustering algorithms led to almost identical patterns of performance for EGA, with Louvain producing marginally better estimates. Given these strong similarities, and to better streamline the commentary of the results, PA_m and PA_{95} will be referred to as PA, *ri*PA_m and *ri*PA_{95} as *ri*PA, EGA_{WT} and EGA_{LV} as EGA, and *ri*EGA_{WT} and *ri*EGA_{LV} as *ri*EGA. Additionally, for each method the MBE and MAE were almost identical across the levels of the independent variables. This implies that when the dimensionality estimates were incorrect, they always erred in the same direction. As the nonzero mean bias error values were positive for all methods, it denotes that the methods erred by suggesting more factors than there were substantive factors in the population.

PLEASE INSERT TABLE 1 ABOUT HERE

The results in Table 1 show that the worst-performing method was PA, followed at a step above by EGA, which proved to be more robust to wording effects than PA but was still considerably impacted. In the case of PA, it maintained almost perfect accuracy for WFL of 0.15, but its accuracy rapidly decreased with higher WFL, to the point of near-zero hit rate levels for 0.30 WFL. Unlike PA, the EGA method maintained almost perfect accuracy for WFL of 0.20, but it decreased noticeably with WLF of 0.25, and sharply for WFL of 0.30. Interestingly, while EGA was completely unaffected by the WFC, the accuracy of the PA method decreased consistently and notably with higher WFC. The MBE criteria revealed that both methods estimated more factors than there were substantive factors in the population and that with the highest WFL of 0.30 they suggested, on average, one more factor than the number of substantive factors. These findings support hypothesis H2. Regarding the proposed random intercept methods, *ri*PA and *ri*EGA improved considerably the performance of PA and EGA, respectively, thus supporting hypothesis H3. However, whereas *ri*EGA maintained excellent levels of accuracy across all levels of WFL and WFC, *ri*PA was only moderately accurate with WFL of 0.30 or WFC of 0.00. In all, *ri*EGA provided the best substantive dimensionality estimates of all the methods considered.

PLEASE INSERT TABLE 2 ABOUT HERE

Table 2 shows the performance of the retention methods across the levels of the four independent variables not related to the wording effects. Regarding the FL, all the methods displayed higher accuracy with higher FL, with the effect being stronger for the traditional formulations of the retention methods. In terms of the FC, they had a negligible impact on all methods except EGA, which provided moderately less accurate dimensionality estimates with higher FC. On the other hand, all the methods except *ri*EGA provided less accurate estimations with more VF, with the effect being stronger, again, for the traditional formulations of the dimensionality methods. Finally, N had a differential effect on PA and *ri*PA compared to EGA and *ri*EGA. Whereas larger levels of N led to less accurate estimates for PA and *ri*PA, they produced more accurate estimates for EGA and *ri*EGA.

To further evaluate the retention methods, ANOVAs were conducted for each method with the MAE as the dependent variable and the six manipulated factors as the independent variables. Up to four-way interactions were estimated in the ANOVAs. The effect sizes for the ANOVAs are shown in Table 3, which includes only those interactions with a medium-sized or larger effect ($\eta_p^2 \geq 0.06$) for at least one of the methods. Of note, *ri*EGA was the *only* factor retention method that did not have an interaction with a large effect size ($\eta_p^2 \geq 0.14$).

PLEASE INSERT FIGURE 2 ABOUT HERE

In the case of PA, the four-way interaction of WFL × WFC × FL × VF achieved a near-large effect, and it subsumed various two-way and three-way interactions with large effect sizes that included the same variables. This four-way interaction is shown in the top panel of Figure 2. The plot in the figure shows that while for some combinations of WFL and WFC the accuracy was equal for all FL levels, for others lower FL led to poorer estimates. This pattern was not uniform, as in some cases differential levels of accuracy across the FL were more pronounced for lower WFC (e.g., WFL = 0.30 & WFC = 0.00) and in other cases for higher WFC (e.g., WFL = 0.20 & WFC = 1.00). In addition, the pattern of differential performance across the FL became more pronounced for higher WFL. Finally, having higher VF led to poorer estimates with decreasing levels of WFC, and this worsening was generally more pronounced for lower FL. It should be noted that when the WFC = 0.00 there were two distinct wording factors, while for WFC = 1.00 there was only one wording factor simulated. With two wording factors, each was defined by fewer variables and was thus more difficult to detect by PA. In this case, having more VF helped to detect both wording factors, which can be seen in the plot as a MAE = 2.00 for the combination of WFL = 0.30, WFC = 0.00, FL = 0.50, and VF = 10. In general, PA only detected one of the two wording factors except when the WFL were highest. On the other hand, when WFC = 1.00 there was only one wording factor, which was easier to detect as it was defined by double the number of variables. Therefore, for WFC = 1.00 the MAE converged to a value of 1.00 (there was an overestimation of one factor) if the WFL were sufficiently high (≥ 0.25).

The results in Table 3 show that *ri*PA produced a similar pattern of salient interactions as PA, with the same interactions obtaining the highest effect sizes for both groups of methods. Thus, for *ri*PA the same four-way interaction of WFL × WFC × FL × VF was interpreted (Figure 2, bottom panel). The interaction plot shows a much more consistent pattern of results for *ri*PA as opposed to PA. For *ri*PA the interaction indicated that having higher WFL led to a differential performance across the FL levels, with lower FL producing the worst estimates, but only for lower WFC, as the dimensionality estimates were all approximately perfect for WFC ≥ 0.75. In addition, having higher VF deteriorated more the performance of structures with lower FL, but mostly only with WFL ≥ 0.25. As can be seen in Figure 2, the conditions where *ri*PA produced the worst results were generally the same as those where PA produced its worst results, indicating that while the random intercept factor helped considerably in approximating the substantive dimensionality, in the most difficult conditions for PA it was not enough.

The EGA method had two large two-way interactions, WFL × FL and WFL × VF, which formed the three-way interaction WFL × FL × VF of medium size. This three-way interaction is shown in Figure 3. The plot in Figure 3 indicates that more VF led to worse dimensionality estimates for EGA, but that this detrimental effect was stronger for lower levels of FL and only occurred for WFL ≥ 0.25. Noteworthy, with WFL = 0.30, VF = 10, and FL = 0.50 the EGA produced its highest MAE of approximately 2.00, which corresponds to detecting the two wording factors.

PLEASE INSERT FIGURE 3 ABOUT HERE

Empirical Example

The data employed in this empirical study was used by Arias et al. (2020), where they evaluated the impact of wording effects on factor modeling. The responses for this publicly available dataset were obtained from 725 U.S. citizens with ages ranging from 18 to 75 years (M = 34.70, SD = 11.70). The participants responded to 18 pairs of Goldberg's (1992) Big Five adjective markers (36 items) corresponding to the dimensions of extraversion, emotional stability, and conscientiousness. Each dimension was balanced, containing the same number of positive (regular) and negative (reversed items). For each adjective that was administered its antonym was also included in the survey, resulting in 18 redundant items. Therefore, for the present study only 18 nonredundant items were analyzed (6 for each dimension). Considering the ordering of the items in the database provided by Arias et al. (2020), for each dimension the *odd* positive items and the *even* negative items were selected for the analyses to prevent item redundancies. The items were responded via a 5-point Likert scale that went from 'very inaccurate' to 'very accurate'. As the item responses were categorical, the dimensionality analyses were conducted on polychoric correlations. The EGA analyses were conducted on the functions *EGA* and *riEGA* of the R package *EGAnet* (version 1.0.1; Golino & Christensen, 2022). Parallel analysis was performed with the function *fa.parallel* from the R package *pysch* (version 2.2.3; Revelle, 2022). The criterion eigenvalues for PA were computed from 1000 data matrices generated from random permutations of the empirical data.

To reproduce our empirical example, we present the code below:

```
# Install latest EGAnet package
if(!"devtools" %in% row.names(installed.packages())){
    install.packages("devtools")}

devtools::install_github("hfgolino/EGAnet")

# Load libraries
library(EGAnet); library(psych); library(foreign); library(ggplot2)

# Load data (Goldberg.sav)
data <- read.spss(
    file.choose(),
    to.data.frame = TRUE)

# Make data numeric (and a matrix object)
data <- simplify2array(lapply(data, as.numeric), higher=FALSE)

# EGA with Walktrap
## Regular EGA
EGAWT <- EGA(data, algorithm="walktrap")

## EGA with random-intercept model
riEGAWT <- riEGA(data, algorithm="walktrap")

# EGA with Louvain
## Regular EGA
EGALV <- EGA(data, algorithm="louvain")

## EGA with random-intercept model
riEGALV <- riEGA(data, algorithm="louvain")

# Parallel analysis
## Eigenvalues for regular PA
PA <- fa.parallel(data, n.iter=1000, sim=FALSE, fa="pc", cor="poly",
                  plot= FALSE, show.legend= FALSE)
Eigen_sample <- PA$pc.values
Eigen_randommean <- apply(PA$values,2,mean)
Eigen_randomperc95 <- apply(PA$values,2,quantile,.95)

## Eigenvalues for PA with random-intercept model
Eigen_residual <- eigen(riEGAWT$RI$correlation)$value

## Eigenvalue plots for regular and random-intercept PA
vars<-c("Sample", "Residual", "Random mean", "Random perc95")
summary_pc<-data.frame(eig =c(Eigen_sample,Eigen_residual,
```

```
                       Eigen_randommean,Eigen_randomperc95),
                       var  = rep(vars,each=length(Eigen_sample)),
                       num = rep(1:length(Eigen_sample),4))
str(summary_pc)
ggplot(summary_pc,aes(x=num,y=eig,colour=var,group=var, shape=var))+
geom_line()+
  geom_point()+
  scale_shape_manual("data",values=c(1,2,3,4),breaks=vars)+
  scale_color_manual("data",values=c("black","darkgrey","grey","lightgrey"),
                     breaks=vars)+
  scale_y_continuous(name='Eigenvalue',breaks=0:max(summary_pc$eig+1)) +
  scale_x_continuous(name='Factor Number',
                     breaks=min(summary_pc$num):max(summary_pc$num)) +
  ggtitle("Parallel Analysis") +
  theme_bw()
```

The dimensionality plots for the EGA methods are shown in Figure 4, while the plots for the PA methods are presented in Figure 5. Regarding the estimates of the traditional EGA methods, the results indicate that both EGA_{WT} and EGA_{LV} suggested four dimensions be retained, rather than the three theoretical dimensions. The two methods, however, differed on the item groupings: in the EGA_{WT} solution the negative items from the extraversion and emotional stability scales were grouped together thus creating an additional dimension, while for EGA_{LV} the added dimension was composed of three negative items, two from extraversion and one from conscientiousness. The dimensionality estimates of the traditional formulations of the PA methods, PA_{95} and PA_m, also suggested that four factors be retained, as the sample eigenvalues corresponding to the fourth factor were higher than both the 95th percentile and mean criterion eigenvalues, while those corresponding to the fifth factor were lower.

PLEASE INSERT FIGURE 4 ABOUT HERE

To obtain the estimates from the proposed methods, a random intercept factor was first estimated. The standardized item loadings on this random intercept factor were 0.27. A residual correlation matrix was obtained by subtracting the model-implied correlation matrix of the random intercept factor from the sample polychoric correlation matrix. The estimates of the "random intercept" methods on this residual correlation matrix suggested that three factors be retained, in line with the theoretical expectations. Both $riEGA_{WT}$ and $riEGA_{LV}$ produced three-factor solutions with all the items grouping according to their theoretical dimension and with each dimension clearly delineated in the network space. Similarly, both *ri*PA95 and *ri*PAm suggested three factors be retained, as the fourth factor eigenvalue from the residual correlation matrix was lower than both the 95^{th} and mean criterion eigenvalues.

PLEASE INSERT FIGURE 5 ABOUT HERE

Discussion

Response biases related to wording effects are ubiquitous for data obtained from mixed-worded psychological scales (Tomás et al., 2013). Although extensive empirical research has shown that wording effects negatively impact latent dimensionality estimates, there is scarce systematic research assessing the problem and no validated solutions had been offered until now. The present study addressed these issues by proposing a substantive dimensionality estimation procedure and systematically testing it through a broad simulation study that employed the EGA and PA retention methods. Additionally, an empirical study was conducted, and a tutorial was offered to exemplify the results of the simulation. Overall, the findings indicate that combining the proposed procedure with both EGA and PA achieves high accuracy in estimating the substantive latent dimensionality underlying the data, but that EGA is superior. Additionally, the present findings shed light into the complex ways that wording effects impact the dimensionality estimates when the response bias in the data is ignored.

Main Findings

Traditional Dimensionality Estimation with Wording Effects

The first objective of this study was to systematically evaluate the impact of wording effects on the dimensionality estimates of recommended factor retention methods. The results from the simulation indicate, as expected, that the retention methods suggest more factors to retain in the presence of wording effects. Thus, with data contaminated by wording effects the dimensionality estimates no longer reflect the underlying substantive dimensionality, but rather, they suggest a combination of the number of substantive and method factors. This finding is consistent with the wording effects empirical factor-analytic literature (García-Batista et al., 2021; Yang et al., 2018; Zhang & Savalei, 2016). However, novel findings of this study show that the impact of the wording effects on the dimensionality estimates is complex, and notably different for the PA and EGA retention methods.

In the case of PA, both the wording factor loadings and wording factor correlations have a strong impact on the dimensionality estimates, and they interact with each other. Because PA is based on the size of the eigenvalues, it will more consistently detect the wording factors when they are better defined (Garrido et al., 2013). This occurs when the wording factor loadings are higher, but also when the wording factor is defined by more variables. The latter is partly a function of the wording factor correlations. With higher wording factors correlations, the multiple wording factors increasingly merge to form a single, larger, wording factor, that is easier to detect for PA. This implies that higher wording factor correlations will lead to more consistent overestimates from PA. However, in the conditions where each wording factor is sufficiently well defined, that is, with higher wording factor loadings and more variables per substantive factor, the overestimation will be greater with lower wording factor correlations, as PA will detect *each* wording factor. Additionally, a fourth key variable that interacts with the three previously discussed, is the substantive factor loadings. Generally, with lower substantive factor loadings PA is better able to detect the wording factors. This finding can be explained through the inherent dependencies of successive eigenvalues (Saccenti et al., 2016). With lower factor loadings, the substantive factors capture a smaller proportion of the total variance in the data, leaving more room for the wording factors to attain higher eigenvalues due to sampling error and capitalization on chance. As a result, the wording factors become more easily detectable by PA.

Regarding the EGA method, its dimensionality estimates were considerably less impacted by the wording effects than PA. Because EGA based on the optimal clustering of variables, and not on the size of the eigenvalues, the substantive structure can remain the optimal choice for lower and moderate levels of wording factor loadings. Nevertheless, with the highest level of wording factor loadings considered here (0.30), EGA usually suggested a dimensionality greater than the number of substantive factors. EGA also overestimated the number of substantive factors with lower substantive factor loadings and more variables per factor, with the two interacting in their impact on EGA. The factor loadings are generally the variable with the greatest influence on the performance of EGA (Golino et al., 2020). Thus, as the wording factor loadings increase, the optimal clustering may shift from being substantially driven to becoming more method driven. Likewise, with more variables per substantive factor, the regular and reversed-coded items may more easily form their own separate clusters, as they are better defined. On the other hand, unlike PA, the wording factor correlations had zero impact on the performance of EGA. This striking difference highlights the complex nature of the impact that wording effects can have on dimensionality estimates across methods.

Substantive Dimensionality Estimation with Wording Effects

The second objective of this study was to propose and test a procedure for substantive dimensionality estimation. Although the increased dimensionality suggested by factor retention methods in the presence of wording effects is not incorrect per se, its usefulness is severely limited because researchers have no clear guidance on what factor model to specify. That is, the dimensionality estimate does not inform if one or more of the suggested dimensions are wording factors, and thus many combinations of substantive + wording factors would be plausible and potentially need to be tested. Also, for dimensionality methods like EGA that inform of the item groupings, wording effects could produce artifactual and uninterpretable solutions. For these reasons, developing a procedure that informs of the substantive dimensionality was thought to be highly beneficial. Our proposed procedure aimed to achieve this goal by estimating a random intercept factor with unit loadings for all the regular and unrecoded reversed items, and then subtracting from the sample correlation the variance implied by this factor.

The findings from the simulation and empirical studies indicate that the proposed random intercept procedure in combination with EGA or PA is highly accurate in estimating the substantive dimensionality underlying the data. This result has several implications. First, it underscores the good performance of the RIIFA strategy in accounting for the wording effects variance, which is line with previous simulation and empirical research (Aichholzer, 2014; Arias et al., 2020; de la Fuente & Abad, 2020; Garrido et al., 2022; Nieto et al., 2021; Savalei & Falk, 2014; Schmalbach et al., 2020; Weydmann et al., 2020). Second, it demonstrates that estimating the random intercept factor separately from the substantive structure can result in a useful approximation of the wording variance in the data. Third, it shows that even if multiple wording factors underlie the observed variables, subtracting from the sample correlation matrix a single random intercept factor is often enough for the dimensionality methods to accurately estimate the substantive dimensionality. Regarding this latter point, the simulations results indicated that for the most difficult conditions of well-defined multiple wording factors (high wording factor loadings, low wording factor correlations, and many variables per substantive factor) the random intercept procedure in conjunction with PA was inaccurate, while in combination with EGA it still performed excellently. As previously discussed, this difference has its main roots in the difference between eigenvalue based (PA) and cluster based (EGA) dimensionality estimation.

Limitations and Future Directions

This study has several limitations that should be considered when assessing its findings. As with any simulation study, the decisions regarding the variables to manipulate and their levels, have an impact on the results. We attempted to perform a broad simulation that manipulated many of the most relevant variables and endeavored to include a range of levels that may be encountered in practice. Nevertheless, more research is necessary to establish the generalizability of our findings. Particularly, in our simulation design the factors composed of mixed-worded items were balanced (same number of regular and reversed items) and their loadings on the method factors were homogeneous. Future research should examine the performance of the proposed procedure for unbalanced scales and unequal method factor loadings. It should be noted, in this regard, that the RIIFA methodology has shown good performance in both of these unfavorable conditions (de la Fuente & Abad, 2020; Savalei & Falk, 2014). Another clear limitation of the proposed procedure is that it assumes that the substantive and wording factors are uncorrelated, something that under some scenarios may not hold. If the substantive and wording factors are correlated to a meaningful degree, the random intercept factor that is subtracted from the sample correlated matrix is likely to contain unwanted substantive variance that could negatively impact the dimensionality assessment. This limitation, nonetheless, is shared with the general RIIFA model, as well as other strategies for modeling method factors, that need orthogonality between the substantive and method factors for model identification (Maydeu & Coffman, 2006).

Practical Implications

The current findings lend support to the use of the proposed random intercept dimensionality procedure for the estimation of the number of substantive factors underlying mixed-worded scales. It is essential that researchers are aware that this procedure is only appropriate for unrecoded data of constructs measured through a combination of regular and reversed items. As shown by the results of the simulation, however, it is not necessary that all constructs are measured by a combination of both types of items. This is particularly important for those cases where there are substantive dimensions underlying the data that had not been anticipated by theory. Additionally, the evidence indicates that researchers can use this procedure as a generalized method for mixed-worded scales. That is, the procedure will accurately estimate the substantive dimensionality of the data even when the amount of wording effects is zero or negligible. This provides an important amount of flexibility in exploratory contexts where the characteristics of the data are less predictable. Also, and particularly for the random intercept EGA, the procedure can accurately estimate the substantive dimensionality even in the presence of multiple method factors in the population.

According to the findings of the simulation, the random intercept procedure led to accurate estimations of the substantive dimensionality for both EGA and PA. However, EGA was superior, providing excellent accuracy in some conditions where PA faltered. Because of this, we consider the random intercept EGA, currently available in the R package *EGAnet*, to be the method of choice for dimensionality estimation with data from mixed-worded scales. In addition, EGA provides an estimate of the item groupings and of the stability of the latent solution (Christensen & Golino, 2021), which make it a highly useful dimensionality assessment method. Of note for both EGA and PA, the variants examined (Walktrap vs. Louvain for EGA and the mean vs. the 95th percentile eigenvalue for PA) produced similar dimensionality estimates, making this selection a choice of minor importance.

Another relevant issue for applied researchers concerns the specification of the factor models after obtaining the dimensionality estimates from the random intercept procedure. When researchers move on to specify an unrestricted factor model it may be necessary to add one or more wording factors. This can be achieved through the exploratory structural equation modeling framework (ESEM; Asparouhov & Muthén, 2009). With ESEM, researchers can specify an unrestricted "exploratory" structure for the substantive factors, and a restricted "confirmatory" one for the wording factors, based on the RIIFA model. To determine if wording factors are necessary, and if they are, how many should be modeled, there are some straightforward guidelines that have been recently proposed (Garrido et al., 2022). For models that include a wording factor *per* substantive factor that is composed of regular and reversed items, researchers may use the item groupings provided by the random intercept EGA to determine which items should load on the different wording method factors.

References

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality*, *53*, 1-4. https://doi.org/10.1016/j.jrp.2014.07.001

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*, 2489-2505. https://doi.org/10.3758/s13428-020-01401-8

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural equation modeling: a multidisciplinary journal*, *16*(3), 397-438. https://doi.org/10.1080/10705510903008204

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

Baumgartner, H., Weijters, B., & Pieters, R. (2018). Misresponse to survey questions: A conceptual framework and empirical test of the effects of reversals, negations, and polar opposite core concepts. *Journal of Marketing Research*, *55*(6), 869-883. https://doi.org/10.1177/0022243718811848

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, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Chan, D. (2010). So why ask me? Are self-report data really that bad?. In *Statistical and methodological myths and urban legends* (pp. 329-356). Routledge.

Christensen, A. P., Garrido, L. E., & Golino, H. (2021). *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. (2021). 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

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155

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

Cosemans, T., Rosseel, Y., & Gelper, S. (2021). Exploratory Graph Analysis for Factor Retention: Simulation Results for Continuous and Binary Data. *Educational and Psychological Measurement*. Advance online publication. https://doi.org/10.1177/00131644211059089

de la Fuente, J., & Abad, F. J. (2020). Comparing methods for modeling acquiescence in multidimensional partially balanced scales. *Psicothema, 32*(4), 590-597. https://doi.org/10.7334/psicothema2020.96

Demetriou, C., Ozer, B. U., and Essau, C. A. (2015). “Self-report questionnaires,” in *The Encyclopedia of Clinical Psychology*, eds R. Cautin and S. Lilienfeld (Malden, MA: John Wiley & Sons, Inc), 1–6. https://doi.org/10.1002/9781118625392.wbecp507

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*(3), 440-464. https://doi.org/10.1207/s15328007sem1303_6

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*, 617–634. https://doi.org/10.1037/met0000167

Ferrando Piera, P. J., Lorenzo Seva, U., Hernández Dorado, A., & Muñiz Fernández, J. (2022). Decálogo para el Análisis Factorial de los Ítems de un Test. *Psicothema*, *34*(1), 7-17. doi: 10.7334/psicothema2021.456

Friedman, J., Hastie, T., & Tibshirani, R. (2014). *glasso: Graphical lasso – estimation of Gaussian graphical models*. Retrieved from https://CRAN.R-project.org/package=glasso

García-Batista, Z. E., Guerra-Peña, K., Garrido, L. E., Cantisano-Guzmán, L. M., Moretti, L., Cano-Vindel, A., Arias, V. B., & Medrano, L. (2021). Using constrained factor mixture analysis to validate mixed-worded psychological scales: The case of the Rosenberg Self-Esteem Scale in the Dominican Republic. *Frontiers in Psychology, 12,* 1-24. https://doi.org/10.3389/fpsyg.2021.636693

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

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

Garrido, L. E., Golino, H., Christensen, A. P., Martínez-Molina, A., Arias, V. B., Guerra-Peña, K., Nieto, M. D., Azevedo, F., & Abad, F. J. (2022). A systematic evaluation of wording effects modeling under the exploratory structural equation modeling framework. https://osf.io/rzu7g/?view_only=cec4abd491104f70943f2412d1ad86ed

Gates, K. M., Henry, T., Steinley, D., & Fair, D. A. (2016). A Monte Carlo evaluation of weighted community detection algorithms. *Frontiers in Neuroinformatics*, *10*, 45. https://doi.org/10.3389/fninf.2016.00045

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Hothorn, T. (2020). mvtnorm: Multivariate Normal and t Distributions. R package version 1.1-1. URL http://CRAN.R-project.org/package=mvtnorm

Gnambs, T., Scharl, A., & Schroeders, U. (2018). The structure of the Rosenberg Self-Esteem Scale: A cross-cultural meta-analysis. *Zeitschrift für Psychologie*, *226*(1), 14–29. https://doi.org/10.1027/2151-2604/a000317

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological Assessment, 4(1), 26-42. https://doi.org/10.1037/1040-3590.4.1.26

Golino, H, & Christensen, A. P. (2022). *EGAnet: Exploratory graph analysis: A framework for estimating the number of dimensions in multivariate data using network psychometrics.* Retrieved from https://CRAN.R-project.org/package=EGAnet

Golino, H. F., & 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. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS one*, *12*(6), 1-16. https://doi.org/10.1371/journal.pone.0174035

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

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

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*, 191-205. https://doi.org/10.1177/1094428104263675

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

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

Juárez-García, A., Merino-Soto, C., Brito-Ortiz, J. F., Nava-Gómez, M. E., & Monroy-Castillo, A. (2021). Is it the perceived stress scale (PSS) Undimimensional and invariant? A Bifactor analysis in Mexican adults. *Current Psychology*, 1-15. Advance online publication. https://doi.org/10.1007/s12144-021-02067-x

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

Kam, C. C. S. (2018). Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociological Methods & Research*, *47*(3), 574-597. https://doi.org/10.1177/0049124115626177

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods*, *48*(3), 936-949. https://doi.org/10.3758/s13428-015-0619-7

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

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

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological methods*, *11*(4), 344-362. https://doi.org/10.1037/1082-989X.11.4.344

Nieto, M. D., Garrido, L. E., Martínez-Molina, A., & Abad, F. J. (2021). Modeling wording effects does not help in recovering uncontaminated person scores: A systematic evaluation with Random Intercept Item Factor Analysis. *Frontiers in Psychology*, *12,* 1-24. https://doi.org/10.3389/fpsyg.2021.685326

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

Pons, P., & Latapy, M. (2005, October). Computing communities in large networks using random walks. In *International symposium on computer and information sciences* (pp. 284-293). Springer, Berlin, Heidelberg.

Revelle, W. (2020). *psych: Procedures for personality and psychological research*. R package version 2.2.3. Retrieved from https://cran.r-project.org/web/packages/psych/index.html

Rosseel Y (2012). “lavaan: An R Package for Structural Equation Modeling.” *Journal of Statistical Software*, 48(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Saccenti, E., & Timmerman, M. E. (2017). Considering Horn’s parallel analysis from a random matrix theory point of view. *Psychometrika*, *82*(1), 186-209. https://doi.org/10.1007/s11336-016-9515-z

Savalei, V., & Falk, C. F. (2014). Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches. *Multivariate behavioral research*, *49*(5), 407-424. https://doi.org/10.1080/00273171.2014.931800

Schmalbach, B., Zenger, M., Michaelides, M. P., Schermelleh-Engel, K., Hinz, A., Körner, A., ... & Brähler, E. (2020). From bi-dimensionality to uni-dimensionality in self-report questionnaires: Applying the random intercept factor analysis model to six psychological tests. *European Journal of Psychological Assessment*, *37*(2), 135-148. https://doi.org/10.1027/1015-5759/a000583

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*(1), 116-131. https://doi.org/10.1509/jmkr.45.1.116

Tomas, J. M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(2), 299-313. https://doi.org/10.1080/10705511.2013.769394

Vigil Colet, A., Navarro González, D., & Morales Vives, F. (2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema, 31*(1), 108-114. https://doi.org/10.7334/psicothema2019.286

Weijters, B., Baumgartner, H., and Schillewaert, N. (2013). Reversed item bias: an integrative model. *Psychological Methods, 18*, 320–334. https://doi.org/10.1037/a0032121

Weydmann, G., Hauck Filho, N., & Bizarro, L. (2020). Acquiescent responding can distort the factor structure of the BIS/BAS scales. *Personality and Individual Differences*, *152*, 1-7. Advance online publication. https://doi.org/10.1016/j.paid.2019.109563

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

Yang, W., Xiong, G., Garrido, L. E., Zhang, J. X., Wang, M. C., & Wang, C. (2018). Factor structure and criterion validity across the full scale and ten short forms of the CES-D among Chinese adolescents. *Psychological assessment*, *30*(9), 1186-1198. https://doi.org/10.1037/pas0000559

Zhang, X., & Savalei, V. (2016). Improving the factor structure of psychological scales: The Expanded format as an alternative to the Likert scale format. *Educational and psychological measurement*, *76*(3), 357-386. https://doi.org/10.1177/0013164415596421

Appendix

Table A1

Proposed Algorithm for Substantive Dimensionality Estimation

```
compute the sample correlation matrix
if the sample correlation matrix is non-positive definite
    smooth the sample correlation matrix
end
compute the random intercept model for the unrecoded dataset
if the estimation of the random intercept model converges
    compute the residual correlation matrix
        if the residual correlation matrix is non-positive definite
            smooth the residual correlation matrix
        end
    compute the dimensionality estimates on the residual correlation matrix
else
    compute the dimensionality estimates on the sample correlation matrix
end
```

Table 1

Performance of the Factor Retention Methods Across the Wording Effects Variables

	WFL				WFC					
Method	0.15	0.20	0.25	0.30	0.00	0.30	0.50	0.70	1.00	Total
					Hit Rate					
PAm	**0.99**	0.77	0.36	0.10	0.74	0.63	0.54	0.47	0.37	0.55
PA95	**0.99**	0.82	0.42	0.13	0.78	0.67	0.58	0.51	0.41	0.59
EGAWT	**1.00**	**0.98**	0.82	0.37	0.79	0.79	0.79	0.79	0.79	0.79
EGALV	**1.00**	**0.98**	0.82	0.36	0.79	0.79	0.79	0.79	0.79	0.79
*ri*PAm	**1.00**	**1.00**	0.94	0.81	0.76	0.93	**0.99**	**1.00**	**1.00**	0.94
*ri*PA95	**1.00**	**1.00**	**0.96**	0.84	0.80	**0.95**	**0.99**	**1.00**	**1.00**	0.95
*ri*EGAWT	**1.00**	**1.00**	**0.99**	0.94	**0.97**	**0.98**	**0.99**	**0.99**	**0.99**	**0.98**
*ri*EGALV	**1.00**	**1.00**	**1.00**	**0.96**	**0.98**	**0.99**	**0.99**	**0.99**	**1.00**	**0.99**
					Mean Bias Error					
PAm	**0.01**	0.23	0.67	1.04	0.42	0.42	0.46	0.53	0.63	0.49
PA95	**0.00**	0.18	0.59	0.98	0.34	0.36	0.42	0.49	0.59	0.44
EGAWT	**0.00**	**0.02**	0.26	1.10	0.35	0.35	0.35	0.35	0.34	0.35
EGALV	**0.00**	**0.02**	0.26	1.10	0.34	0.35	0.35	0.35	0.34	0.35
*ri*PAm	**0.00**	**0.00**	**0.06**	0.19	0.24	**0.07**	**0.01**	**0.00**	**0.00**	**0.06**
*ri*PA95	**0.00**	**0.00**	**0.04**	0.15	0.19	**0.04**	**0.00**	**0.00**	**0.00**	**0.05**
*ri*EGAWT	**0.00**	**0.00**	**0.01**	**0.07**	**0.04**	**0.02**	**0.02**	**0.01**	**0.01**	**0.02**
*ri*EGALV	**0.00**	**0.00**	**0.00**	**0.05**	**0.03**	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**
					Mean Absolute Error					
PAm	**0.01**	0.23	0.67	1.04	0.42	0.42	0.46	0.53	0.63	0.49
PA95	**0.01**	0.18	0.60	0.98	0.35	0.36	0.42	0.49	0.59	0.44
EGAWT	**0.00**	**0.02**	0.26	1.10	0.35	0.35	0.35	0.35	0.34	0.35
EGALV	**0.00**	**0.02**	0.26	1.10	0.34	0.35	0.35	0.35	0.34	0.35
*ri*PAm	**0.00**	**0.00**	**0.06**	0.19	0.24	**0.07**	**0.01**	**0.00**	**0.00**	**0.06**
*ri*PA95	**0.00**	**0.00**	**0.04**	0.16	0.20	**0.05**	**0.01**	**0.00**	**0.00**	**0.05**
*ri*EGAWT	**0.00**	**0.00**	**0.01**	**0.07**	**0.04**	**0.02**	**0.02**	**0.01**	**0.01**	**0.02**
*ri*EGALV	**0.00**	**0.00**	**0.00**	**0.05**	**0.03**	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**

Note. WFL = wording factor loadings; WFC = wording factor correlations; PAm = parallel analysis with the mean criterion; PA95 = parallel analysis with the 95th percentile criterion; EGAWT = exploratory graph analysis with the Walktrap algorithm; EGALV = exploratory graph analysis with the Louvain algorithm; *ri* = random intercept. Cells highlighted in light grey correspond to hit ratios between 0.90 and 0.94, mean bias errors between 0.11 and 0.20, and mean absolute errors between 0.11 and 0.20. Cells bolded and highlighted in dark grey correspond to hit ratios between 0.95 and 1.00, mean bias errors between 0.00 and 0.10, and mean absolute errors between 0.00 and 0.10.

Table 2

Performance of the Factor Retention Methods Across the Remaining Independent Variables

	FL			FC			VF		N		
Method	0.50	0.60	0.70	0.00	0.30	0.50	6	10	300	500	1000
					Hit Rate						
PA_m	0.41	0.56	0.69	0.55	0.55	0.56	0.66	0.45	0.61	0.55	0.50
PA_{95}	0.46	0.60	0.72	0.59	0.59	0.59	0.70	0.48	0.66	0.59	0.53
EGA_{WT}	0.66	0.84	0.88	0.86	0.79	0.73	0.88	0.70	0.76	0.79	0.83
EGA_{LV}	0.65	0.84	0.88	0.86	0.79	0.72	0.88	0.70	0.76	0.78	0.83
riPA_m	0.89	0.94	**0.98**	0.94	0.94	0.94	**0.97**	0.90	**0.95**	0.94	0.92
riPA_{95}	0.91	**0.95**	**0.98**	**0.95**	0.95	0.94	**0.98**	0.92	**0.96**	**0.95**	0.93
riEGA_{WT}	**0.95**	**1.00**	**1.00**	**0.99**	**0.99**	**0.97**	**0.98**	**0.99**	**0.97**	**0.98**	**0.99**
riEGA_{LV}	**0.97**	**1.00**	**1.00**	**0.99**	**0.99**	**0.98**	**0.99**	**0.99**	**0.98**	**0.99**	**1.00**
					Mean Bias Error						
PA_m	0.67	0.48	0.32	0.49	0.49	0.48	0.36	0.62	0.42	0.49	0.56
PA_{95}	0.60	0.43	0.29	0.45	0.44	0.43	0.30	0.57	0.35	0.44	0.53
EGA_{WT}	0.59	0.25	0.20	0.22	0.35	0.47	0.18	0.52	0.39	0.36	0.30
EGA_{LV}	0.59	0.25	0.20	0.22	0.35	0.46	0.18	0.52	0.38	0.36	0.30
riPA_m	0.11	**0.06**	**0.02**	**0.06**	**0.06**	**0.06**	**0.03**	**0.10**	**0.04**	**0.06**	**0.08**
riPA_{95}	**0.08**	**0.04**	**0.02**	**0.05**	**0.05**	**0.04**	**0.02**	**0.08**	**0.02**	**0.05**	**0.07**
riEGA_{WT}	**0.06**	**0.00**	**0.00**	**0.01**	**0.02**	**0.03**	**0.02**	**0.02**	**0.03**	**0.02**	**0.01**
riEGA_{LV}	**0.04**	**0.00**	**0.00**	**0.01**	**0.01**	**0.02**	**0.01**	**0.01**	**0.03**	**0.02**	**0.00**
					Mean Absolute Error						
PA_m	0.67	0.48	0.32	0.49	0.49	0.48	0.36	0.62	0.42	0.49	0.56
PA_{95}	0.61	0.43	0.29	0.45	0.44	0.44	0.31	0.57	0.36	0.44	0.53
EGA_{WT}	0.59	0.25	0.20	0.22	0.35	0.47	0.18	0.52	0.39	0.36	0.30
EGA_{LV}	0.59	0.25	0.20	0.22	0.35	0.46	0.18	0.52	0.38	0.36	0.30
riPA_m	0.11	**0.06**	**0.02**	**0.06**	**0.06**	**0.06**	**0.03**	**0.10**	**0.05**	**0.06**	**0.08**
riPA_{95}	**0.10**	**0.05**	**0.02**	**0.05**	**0.05**	**0.06**	**0.03**	**0.08**	**0.04**	**0.05**	**0.07**
riEGA_{WT}	**0.06**	**0.00**	**0.00**	**0.01**	**0.02**	**0.04**	**0.02**	**0.02**	**0.03**	**0.02**	**0.01**
riEGA_{LV}	**0.04**	**0.00**	**0.00**	**0.01**	**0.01**	**0.02**	**0.01**	**0.01**	**0.03**	**0.02**	**0.00**

Note. FL = substantive factor loadings; FC = substantive factor correlations; VF = variables per factor; N = sample size; PA_m = parallel analysis with the mean criterion; PA_{95} = parallel analysis with the 95th percentile criterion; EGA_{WT} = exploratory graph analysis with the Walktrap algorithm; EGA_{LV} = exploratory graph analysis with the Louvain algorithm; *ri* = random intercept. Cells highlighted in light grey correspond to hit ratios between 0.90 and 0.94, mean bias errors between 0.11 and 0.20, and mean absolute errors between 0.11 and 0.20. Cells bolded and highlighted in dark grey correspond to hit ratios between 0.95 and 1.00, mean bias errors between 0.00 and 0.10, and mean absolute errors between 0.00 and 0.10.

Table 3

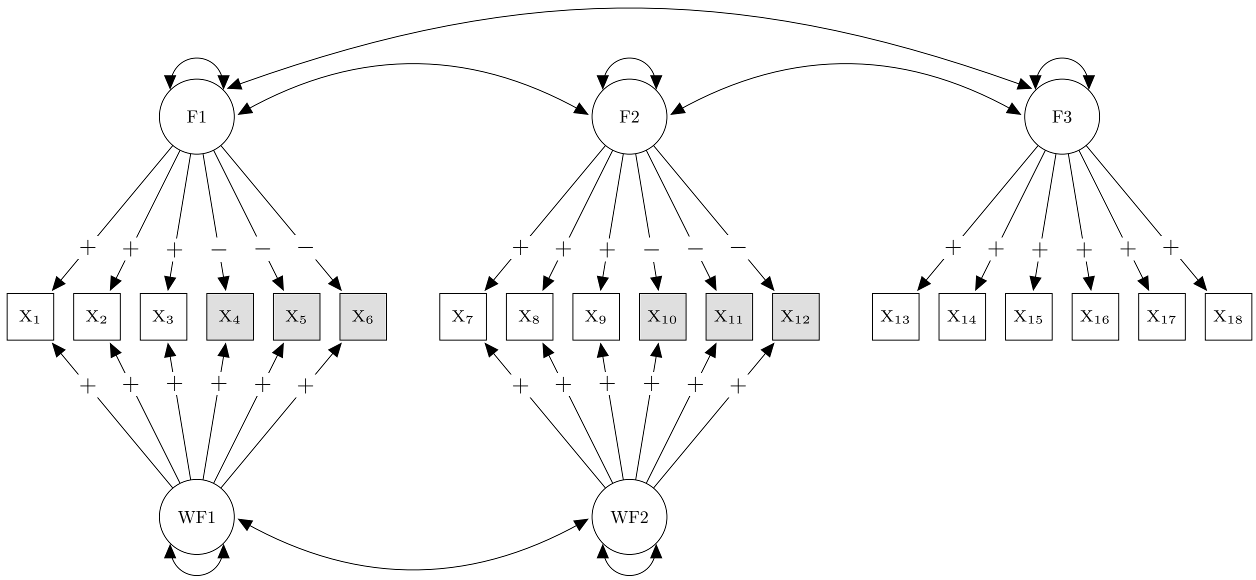
ANOVA Effect Sizes for the Mean Absolute Error Dependent Variable

Effect	PA_m	PA_{95}	EGA_{WT}	EGA_{LV}	*ri*PA_m	*ri*PA_{95}	*ri*EGA_{WT}	*ri*EGA_{LV}
Main effects								
WFL	**0.72**	**0.71**	**0.65**	**0.67**	**0.26**	**0.20**	0.04	0.03
WFC	0.10	0.12	0.00	0.00	**0.33**	**0.27**	0.01	0.01
FL	**0.25**	**0.22**	**0.22**	**0.23**	0.08	0.06	0.03	0.03
FC	0.00	0.00	0.09	0.09	0.00	0.00	0.01	0.00
VF	**0.22**	**0.23**	**0.21**	**0.22**	0.06	0.04	0.00	0.00
N	0.05	0.07	0.01	0.01	0.01	0.01	0.00	0.01
Two-way interactions								
WFL × FC	0.00	0.00	0.07	0.08	0.00	0.00	0.01	0.00
WFC × FL	0.04	0.03	0.00	0.00	0.11	0.09	0.01	0.01
WFL × FL	0.09	0.09	**0.19**	**0.20**	0.08	0.06	0.07	0.06
WFC × VF	0.03	0.03	0.00	0.00	0.09	0.09	0.00	0.00
WFL × VF	0.09	0.11	**0.33**	**0.34**	0.07	0.07	0.00	0.00
WFL × WFC	**0.16**	**0.14**	0.00	0.00	**0.39**	**0.35**	0.01	0.01
Three-way interactions								
WFL × FL × FC	0.00	0.00	0.08	0.08	0.00	0.00	0.02	0.00
WFL × FL × VF	0.03	0.02	0.09	0.10	0.01	0.01	0.00	0.00
WFL × WFC × FL	**0.16**	0.13	0.01	0.01	0.12	0.10	0.02	0.02
WFL × WFC × VF	**0.15**	**0.17**	0.00	0.00	0.10	0.11	0.00	0.00
Four-way interactions								
WFL × WFC × FL × VF	0.13	0.11	0.00	0.00	0.10	0.08	0.00	0.00

Note. WFL = wording factor loadings; WFC = wording factor correlations; FL = substantive factor loadings; FC = substantive factor correlations; VF = variables per factor; N = sample size; PA_m = parallel analysis with the mean criterion; PA_{95} = parallel analysis with the 95th percentile criterion; EGA_{WT} = exploratory graph analysis with the Walktrap algorithm; EGA_{LV} = exploratory graph analysis with the Louvain algorithm; *ri* = random intercept. Cells bolded and highlighted in dark grey indicate large partial eta squared effect sizes (≥ 0.14).

Figure 1

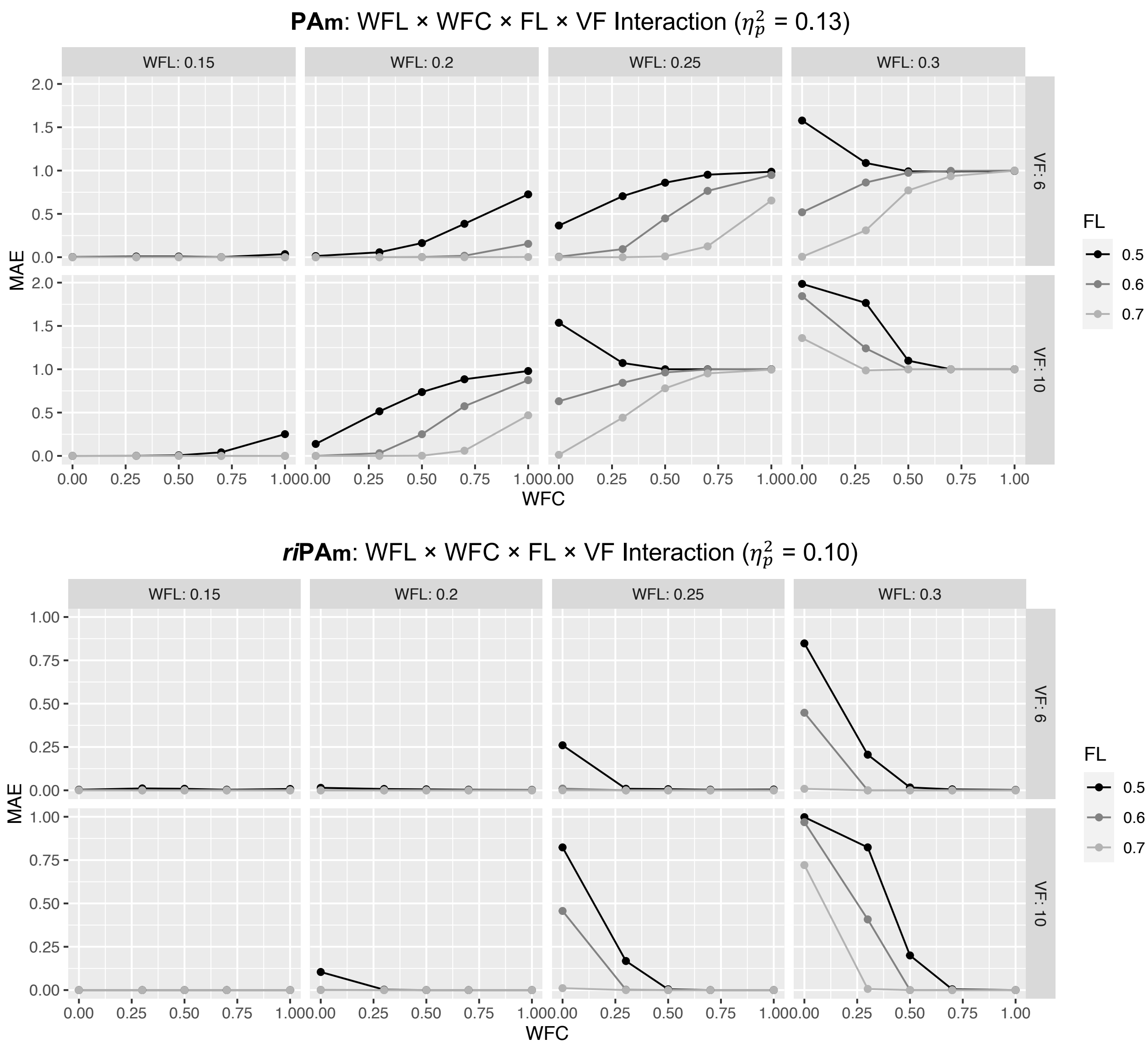
Data Generating Model



Note. F1-F3 = substantive factors; WF1-WF2 = wording factors; X_1-X_{18} = items. Squares filled in grey indicate unrecoded reversed items. Unidirectional arrows linking circles and rectangles represent the factor loadings. Bidirectional arrows linking the circles represent the factor covariances/correlations. Bidirectional arrows connecting a single circle represent the factor variances. For simplicity, the item uniquenesses have been omitted. The number of items per factor had two levels, 6 (as depicted in the plot) and 10.

Figure 2

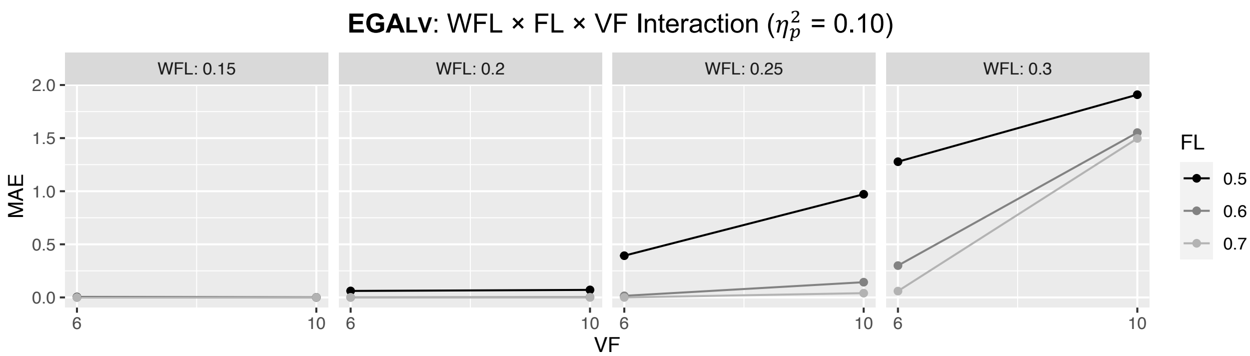
ANOVA Interactions for PAm and riPAm with the MAE Dependent Variable



Note. PAm = parallel analysis with the mean criterion; *ri*PAm = random intercept parallel analysis with the mean criterion; MAE = mean absolute error; WFL = wording factor loadings; WFC = wording factor correlations; FL = factor loadings; VF = variables per factor; η_p^2 = partial eta squared effect size.

Figure 3

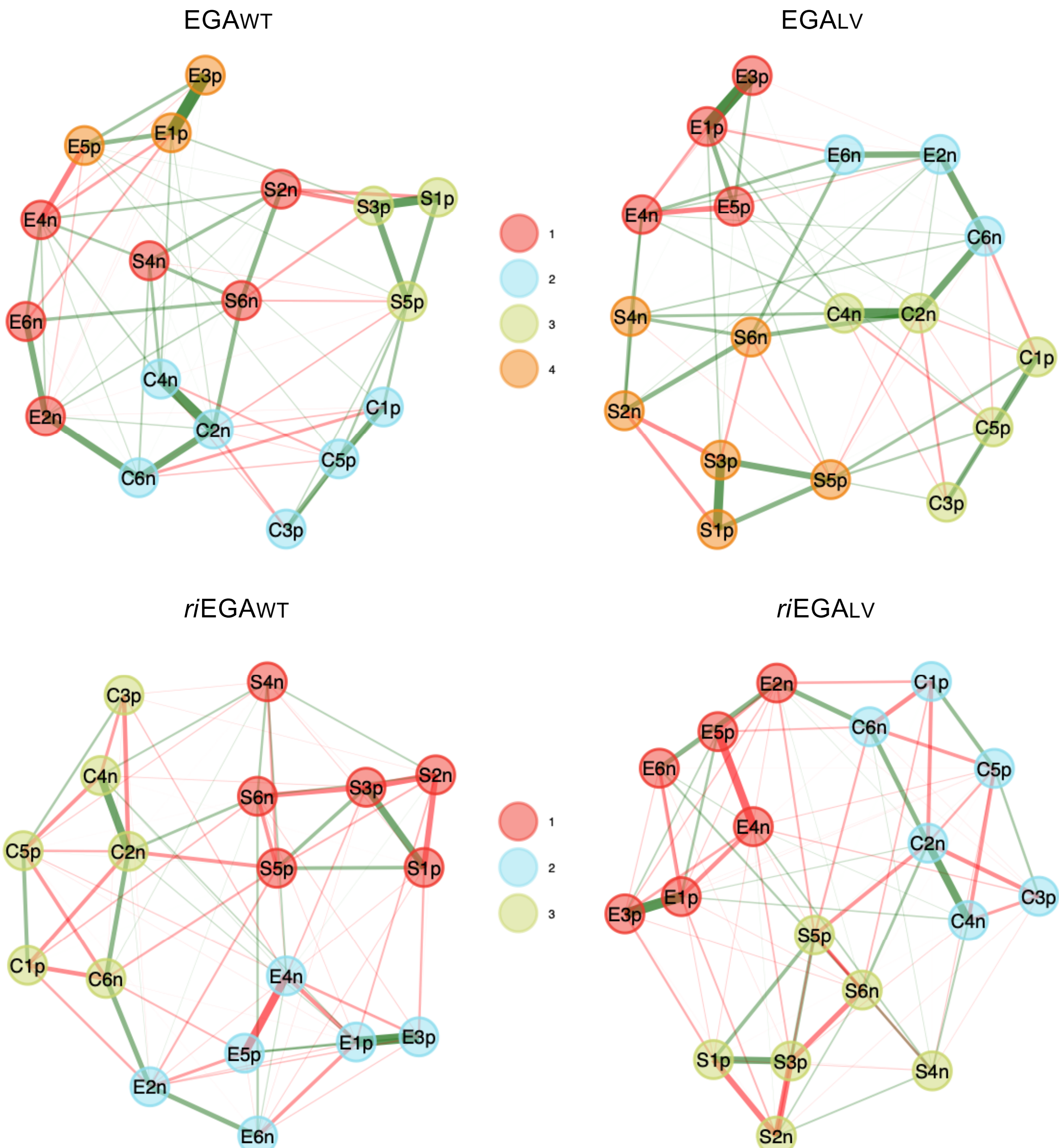
ANOVA Interaction for EGALV with the MAE Dependent Variable



Note. EGALV = exploratory graph analysis with the Louvain algorithm; MAE = mean absolute error; WFL = wording factor loadings; FL = factor loadings; VF = variables per factor; η_p^2 = partial eta squared effect size.

Figure 4

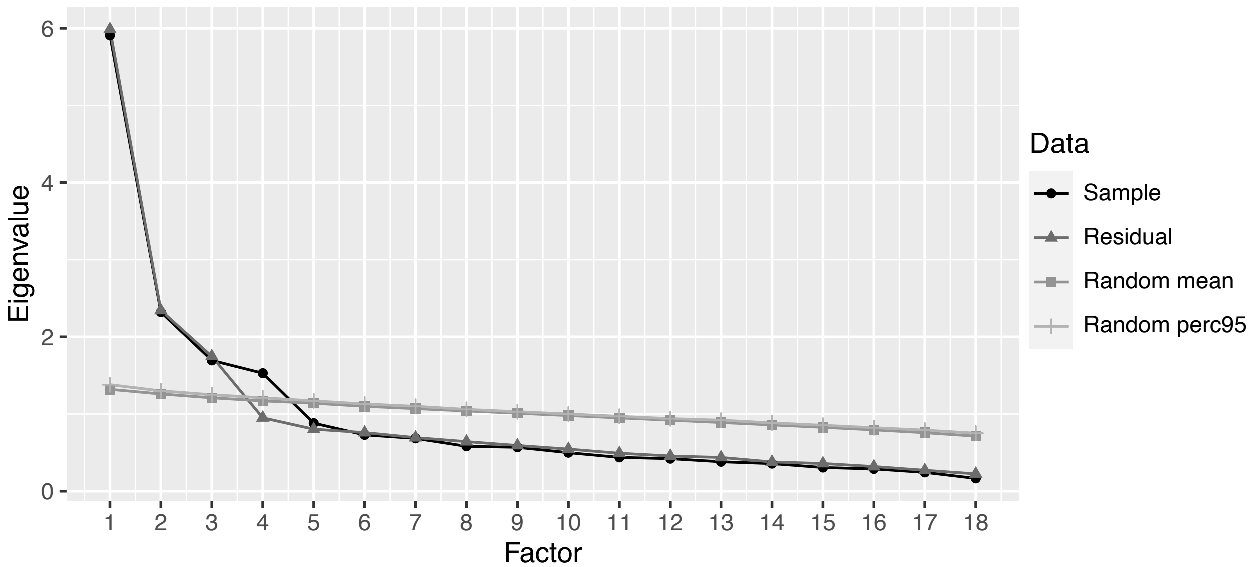
Exploratory Graph Analysis Dimensionality Estimates for the Empirical Data



Note. EGAWT = exploratory graph analysis with the Walktrap algorithm; EGALV = exploratory graph analysis with the Louvain algorithm; *ri* = random intercept; E1p = extraverted; E2n = unenergetic; E3p = talkative; E4n = timid; E5p = assertive; E6n = unadventurous; S1p = calm; S2n = tense; S3p = at ease; S4n = envious; S5p = stable; S6n = discontented; C1p = organized; C2n = irresponsible; C3p = conscientiousness; C4n impractical; C5p = thorough; C6n = lazy; The E, S, and C in the item names denote the dimensions of extraversion, emotional stability, and conscientiousness, respectively. The p after the item number denotes a positive (regular) item. The n after the item number denotes a negative (reversed) item. The reversed items were left unrecoded for the analyses.

Figure 5

Parallel Analysis Dimensionality Estimates for the Empirical Data



Note. Sample = eigenvalues derived from the sample correlation matrix; Residual = eigenvalues derived from the residual correlation matrix obtained after subtracting from the sample correlation matrix the model-implied correlation matrix of the random intercept factor; Random mean = mean eigenvalues computed from 1000 datasets obtained through random permutations of the sample data; Random perc95 = 95th percentile eigenvalues computed from 1000 datasets obtained through random permutations of the sample data.