

Entropy Fit Indices: New Fit Measures for Assessing the Structure and Dimensionality of Multiple Latent Variables

Hudson Golino¹, Robert Moulder¹, Dingjing Shi¹, Alexander P. Christensen², Luis Eduardo Garrido³, Maria Dolores Nieto⁴, John Nesselrode¹, Ritu Sadana⁵, Jotheeswaran Amuthavalli Thiyagarajan⁵, & Steven M. Boker¹

¹Department of Psychology, University of Virginia; ²Department of Psychology, University of North Carolina at Greensboro; ³Department of Psychology, Pontificia Universidad Catolica Madre y Maestra; ⁴Department of Psychology, Universidad Antonio de Nebrija; ⁵Ageing and Health Unit, Department of Maternal, Newborn, Child, Adolescent Health and Ageing, World Health Organization

ARTICLE HISTORY

Compiled June 3, 2020

ABSTRACT

The accurate identification of the content and number of latent factors underlying multivariate data is an important endeavor in many areas of Psychology and related fields. Recently, a new dimensionality assessment technique based on network psychometrics was proposed (Exploratory Graph Analysis, EGA), but a measure to check the fit of the dimensionality structure to the data estimated via EGA is still lacking. Although traditional factor-analytic fit measures are widespread, recent research has identified limitations for their effectiveness in categorical variables. Here, we propose three new fit measures (termed entropy fit indices) that combines information theory, quantum information theory and structural analysis: Entropy Fit Index (EFI), EFI with Von Neuman Entropy (EFI.vn) and Total EFI.vn (TEFI.vn). The first can be estimated in complete datasets using Shannon entropy, while EFI.vn and TEFI.vn can be estimated in correlation matrices using quantum information metrics. We show, through several simulations, that TEFI.vn, EFI.vn and EFI are as accurate or more accurate than traditional fit measures when identifying the number of simulated latent factors. However, in conditions where more factors are extracted than the number of factors simulated, only TEFI.vn presents a very high accuracy. In addition, we provide an applied example that demonstrates how the new fit measures can be used with a real-world dataset, using exploratory graph analysis.

KEYWORDS

Fit measures; Dimensionality Analysis; Information Theory; Network Psychometrics; Exploratory Graph Analysis

Introduction

Estimating the dimensionality or the number of underlying dimensions (or factors) in multivariate datasets is an important step in psychology and related areas. In the health and social sciences areas, assessing the dimensionality of scales, tests, and questionnaires is one of the first steps on the path to understanding multivariate data. Parallel, exploratory factor, and principal component analysis (e.g., eigenvalues and scree plots) are the most common techniques implemented in commercial software and among the most common strategies used to identify the dimensionality of the data across research areas. Despite their long history and widespread application, simulation studies have consistently shown that traditional factor-analytic techniques present important limitations (Crawford et al., 2010; Garrido, Abad, & Ponsoda, 2013; Hayton, Allen, & Scarpello, 2004; Keith, Caemmerer, & Reynolds, 2016; Ruscio & Roche, 2012).

Recently, a new dimensionality assessment technique termed *Exploratory Graph Analysis* (**EGA**: Golino & Epskamp, 2017) was proposed. EGA has shown a number of advantages when compared to traditional factor-analytic techniques (Golino & Demetriou, 2017; Golino & Epskamp, 2017). The technique estimates the number of underlying dimensions by first modeling multivariate data as a network of partial correlations (using the *Gaussian graphical model*, **GGM**; Lauritzen, 1996) or zero-order correlations (using the *triangulated maximally filtered graph* approach, **TMFG**; Massara, Di Matteo, & Aste, 2016). Then, the technique applies a community detection algorithm for weighted networks (i.e., Walktrap; Pons & Latapy, 2005), which identifies the number of dimensions in the network using a random walk process. Not only does EGA present higher accuracy than many traditional methods (i.e., percentage of correct estimates across different simulated conditions), but it also produces a visual guide (network plot) that indicates the number of dimensions to retain, showing which items cluster together and their level of association, all without the need of any specification from researchers (e.g., type of rotation, number of factors; Golino et al., 2020). Although useful, EGA lacks a fit index to compare different dimensionality structures between EGA methods (**GGM** and **TMFG**), EGA and factor-analytic methods, and EGA and theoretical structures. Moreover, community detection algorithms often have parameters (e.g., steps in the Walktrap algorithm) that adjust the number of dimensions that the algorithm will find. Therefore, a fit measure that can be used to automatically adjust these parameters for optimal performance could greatly improve EGA's precision.

Applied researchers and methodologists use fit indices to assess the factor solutions of their data (or instruments), mainly because they provide useful diagnostic information, but also because they enable the comparison of different dimensionality structures (i.e., number of factors and distribution of items per factors). Despite the usefulness of fit indices used in factor analysis to estimate the number of underlying dimensions, two main limitations can be found in the literature. First, there are only a few simulation studies investigating the accuracy of the traditional fit indices (such as Comparative Fit Index and Root Mean Square Error of Approximation or CFI and RMSEA, respectively) to estimate the correct number of factors (Barendse, Oort, & Timmerman, 2015; Clark & Bowles, 2018; Frazier & Youngstrom, 2007; Garrido, Abad, & Ponsoda, 2016; Yang & Xia, 2015). Second, they produce poor dimensionality estimates for skewed categorical variables (Clark & Bowles, 2018; Garrido et al., 2016). Clearly, there is a need for fit indices that mitigate these limitations.

The goals of the paper are twofold. First, we introduce a new family of fit indices termed *Entropy Fit Indices (EFIs)* will be presented: an index that is computed using the raw data (*Entropy Fit Index; EFI*), and two indices that are computed using the correlation matrix (*Entropy Fit Index with Von Neumann Entropy* and *Total Entropy Fit Index; vnEFI* and *TEFI*, respectively). These new fit indices can be used to assess the degree of disorder (or uncertainty, using information theory terms) of a set of random variables for structures with different skew of variables, number of factors, different distribution of items in factors, and placement of items in each factor. Lower values of the EFIs indicate lower disorder (or uncertainty) of a system of variables, indicating a higher probability that a given structure represents the best organization of these variables. Second, we use a Monte Carlo simulation method to compare the EFIs to some of the most commonly used fit indices in factor analysis: CFI, RMSEA, Tucker-Lewis Index (TLI), and Standardized Root Mean Square Residual (SRMR).

This article is organized as follows: The first section briefly introduces several traditional factor-analytic fit indices and examines their usefulness in estimating the number of underlying dimensions in multivariate data. The second section briefly reviews EGA, and summarizes the main findings of previous simulation studies. The third section briefly introduces the concept of entropy, its properties and estimation methods, and the new EFIs. The fourth section presents a Monte Carlo simulation, where the number of factors, variables per factor, interfactor correlations, factor loadings, and skewness of the variables are systematically controlled. The fit indices will be computed when a given structure reflects: (1) exactly the simulated factor structure, (2) the correct number of factors, but with a random placement of items per factor, and (3) an incorrect number of factors (underfactoring or overfactoring). The fifth section uses an empirical example, from a large scale study on aging, to demonstrate the use of EFI. Finally, the last section presents a discussion of the results, limitations of the current study, and future directions.

Fit Indices used in Dimensionality Assessment

Fit indices derived from factor analysis models have been proposed as tools to estimate latent dimensions in multivariate data (Santos, Vagos, & Rijo, 2018; Schermelleh-Engel, Moosbrugger, & Müller, 2003; Ventimiglia & MacDonald, 2012). These fit indices include (among others): CFI (Bentler, 1990), TLI (Tucker & Lewis, 1973), RMSEA (Steiger, 1980), and SRMR (Joreskog & Sorbom, 1981). The first three indices are based on the noncentrality parameter (λ_M) of the specified model, which is computed as $\chi_M^2 - df_M$, where χ_M^2 is the chi-square statistic that tests the equivalence of the population covariance matrix of observed variables and the model-implied covariance matrix, and df_M is the degrees of freedom of the specified model. The last fit index, SRMR, is based on the average difference between observed and model-implied covariance matrices. The CFI and TLI indices are considered incremental fit indices because they evaluate the degree to which the specified model is superior to a baseline model—usually a null model where the observed variables are proposed to be uncorrelated—in reproducing the observed covariances. In contrast, the RMSEA and SRMR are considered absolute fit indices because they offer a measure of overall fit that only takes into account the fit of the specified model. A key advantage of the RMSEA index is that, because it follows a known χ^2 distribution asymptotically, confidence intervals can be estimated around the point estimate, allowing for formal hypothesis

tests to be conducted by jointly considering the point estimate and its associated confidence interval (Chen et al., 2008). The formulas for the aforementioned fit indices are shown below in Equations 1 to 4:

$$CFI = 1 - \frac{\max(\lambda_M, 0)}{\max(\lambda_N, \lambda_M)}, \quad (1)$$

$$TLI = 1 - \left(\frac{\lambda_M}{\lambda_N} \right) \left(\frac{df_N}{df_M} \right), \quad (2)$$

$$RMSEA = \max \left(\sqrt{\frac{\lambda_M}{df_M(N-1)}}, 0 \right), \quad (3)$$

$$SRMR = \sqrt{\frac{\sum_{i=1}^p \sum_{j=1}^i \left(\frac{s_{ij}}{\sqrt{s_{ii}\sqrt{s_{jj}}} - \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\sqrt{\hat{\sigma}_{jj}}}} \right)^2}{\frac{p(p+1)}{2}}}, \quad (4)$$

where λ_N and df_N are the noncentrality parameter and degrees of freedom of the baseline model, N is the sample size, s_{ij} is the observed covariance, $\hat{\sigma}_{ij}$ is the model-implied covariance, s_{ii} and s_{jj} are the observed standard deviations, $\hat{\sigma}_{ii}$ and $\hat{\sigma}_{jj}$ are the model-implied standard deviations, and p is the number of observed variables.

SRMR (Equation 4) and the fit index for categorical data termed weighted root mean square residual (WRMR) are residual based fit indices (DiStefano, Liu, Jiang, & Shi, 2018). The WRMR fit index was originally proposed by Muthén and Muthén (1998) and first examined by Yu (2002). Among several fit indices evaluated (e.g., CFI, TLI, RMSEA), WRMR produced results that were most similar to the SRMR index (Yu, 2002). Initially the WRMR was recommended for models composed of variables that have widely disparate variances and/or are on different scales (Yu, 2002).

There are several reasons why we opted to use the SRMR over WRMR for the present study. First, the index has been regarded as “experimental” by its creators (Muthén and Muthén, 1998-2012). Second, in contrast of other well-established fit indices it has only been evaluated in a couple of simulation studies (DiStefano et al., 2018). Third, and most importantly, based on the results of DiStefano et al. (2018) the creators of the index decided to remove it from the output of the MPlus program in favor of the SRMR index (Asparouhov & Muthén, 2018). The authors state “In Mplus 8.1 for the WLS family of estimators this SRMR fit index replaces the WRMR fit index which has been shown to perform poorly for situations with extremely large sample sizes, see DiStefano et al. (2018)” (Asparouhov and Muthén (,2018); p.11). Finally, all the simulated items were on the same scale.

When using fit indices to estimate latent dimensionality under the framework of unrestricted (exploratory) factor models, the common rationale is that factors should be sequentially added to a model until an acceptable level of fit has been achieved. First, a one-factor model would be fitted to the data and its level of fit would be compared against pre-specified cut-off values for the fit indices. If the fit was deemed acceptable, the one-factor model would be retained; if not, an unrestricted two-factor model would be estimated and the process would be repeated, until an acceptable level of fit was achieved. Applied researchers that have employed this approach (e.g., Campbell-Sills, Liverant, & Brown, 2004; Sanne, Torsheim, Heiervang, & Stormark, 2009) have generally relied on the typical cut-off values that have been proposed for these indices in the literature, such as .90 or .95 for CFI and TLI, and .05 or .08 for RMSEA and SRMR (good and acceptable fit, respectively; Browne & Cudeck, 1992; Chen et al., 2008; Hu & Bentler, 1999a).

Despite the appeal of using fit indices to estimate latent dimensionality, recent simulation studies have shown that they are not the best tools for this endeavor. Garrido et al. (2016) examined the performance of these four fit indices across a large number of conditions of continuous and categorical variables and using a wide range of cut-off values for each index. They found that the CFI and TLI indices were the most accurate, followed at a step below by RMSEA, and finally by SRMR, which produced very poor estimates across all the cut-off values in its relevant range of functioning. Even the best performing fit indices (CFI and TLI) were more variable and less accurate than parallel analysis—one of the most widely recommended methods in the dimensionality literature (Garrido et al., 2016; Horn, 1965). Subsequent studies by Clark and Bowles (2018) and Beierl, Bühner, and Heene (2018) have corroborated these findings, showing that none of the conventional cut-off values previously cited appeared to perform well enough to be recommended. Similar to Garrido et al. (2016), they found that the RMSEA index combined with the typical cut-off value of 0.05 was insensitive to latent misspecification, frequently accepting models with fewer major factors than those in the population. Similarly, while the CFI and TLI indices performed more reliably using a 0.95 cut-off value, their accuracy was contingent on the size of the factor loadings, factor correlations, and number of items (Beierl et al., 2018; Clark & Bowles, 2018). Again, the SRMR index performed very poorly, generally accepting underfactored models when using a cut-off value of 0.08.

In general, research studies evaluating the use of fit indices for dimensionality assessment have encountered similar problems to those found in the general literature that has assessed the usefulness of fit indices for latent variable modeling. More specifically, it has been shown that the values of the fit indices are affected by incidental parameters not related to the size of the misfit, such as the size of the factor loadings, the number of response categories, and the sample size (Chen et al., 2008; Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011; Hu & Bentler, 1999a; McNeish, An, & Hancock, 2018; Savalei, 2012; Xia & Yang, 2018). Because of this, there are no “golden rules” for fit indices as no single cut-off value can be expected to perform well across the diverse type of data structures encountered in practice (Marsh, Hau, & Wen, 2004). In a recent study, Clark and Bowles (2018) concluded that, “the common fit statistics cut-offs employed in the literature are at best imperfect tools for guiding decisions regarding dimensionality when interpreting an item factor analysis” (Clark & Bowles, 2018, p. 555). Clark and Bowles (2018) suggest that fit indices should be used in conjunction with other methods for assessing dimensionality.

Exploratory Graph Analysis: An Overview

A recently developed dimensionality assessment technique, Exploratory Graph Analysis (Golino & Epskamp, 2017), is part of a relatively new field of quantitative methods termed *network psychometrics* (Epskamp, 2018; Epskamp, Rhemtulla, & Borsboom, 2017). EGA estimates the dimensionality structure of multivariate data by combining network analysis with a community detection algorithm (i.e., a method to detect dimensions in networks), and has shown to be a promising dimensionality assessment technique in simulation studies (Golino & Epskamp, 2017; Golino et al., 2020). Currently, however, there is no fit index that can directly determine the adequacy of the dimensional structures estimated via EGA (without having to use additional models such as CFA; Golino & Demetriou, 2017). This section briefly introduces EGA, which can be used as one of the steps for assessing dimensionality, complementing the suggestion of Clark and Bowles (2018).

EGA is a dimensionality technique from the network psychometrics perspective. A network is depicted by nodes (circles) representing variables and edges (lines) representing the associations between variables. The general notion behind the EGA algorithm is to estimate a network and apply the Walktrap community detection algorithm (Pons & Latapy, 2005), which is a common approach for identifying *communities* (densely connected sets of nodes) in weighted networks. These communities are shown to be statistically equivalent to latent factors of factor models (Golino & Epskamp, 2017; Golino et al., 2020). Below we provide a more thorough explanation of this algorithm. Readers interested in a more in-depth description of EGA can find additional details in the simulation study and tutorial published by Golino et al. (2020).

The first network estimation method for EGA algorithm is the Gaussian graphical model (GGM; Lauritzen, 1996), which estimates multivariate joint probability distribution of random variables—that is, the probability for the values of each pair of variables to be within the range of values for those variables. This probability distribution is represented by the inverse of the covariance matrix (often referred to as the *precision matrix*), whose off-diagonal elements represent the conditional dependence between two variables given all other variables. The current approach used in network psychometrics to estimate a GGM works as follows: Let’s assume a set of random variables \mathbf{y} that are normally distributed with mean zero and variance-covariance matrix $\mathbf{\Sigma}$, and let \mathbf{K} (kappa) be the inverse of $\mathbf{\Sigma}$,

$$\mathbf{K} = \mathbf{\Sigma}^{-1} \tag{5}$$

then, each element k_{ij} (row i , column j of \mathbf{K}) can be standardized to yield the partial correlation between two variables y_i and y_j , given all other variables in \mathbf{y} (Epskamp & Fried, 2018):

$$\text{Cor}(Y_i, Y_j | \mathbf{y}_{-(i,j)}) = -\frac{k_{ij}}{\sqrt{k_{ii}}\sqrt{k_{jj}}} \tag{6}$$

Epskamp and Fried (2018) pointed out that modeling \mathbf{K} in a way that every nonzero element is treated as a freely estimated parameter generates a sparse model for Σ . The sparse model of the variance-covariance matrix is the GGM (Epskamp & Fried, 2018).

This partial correlation matrix represents a fully connected network where all nodes are connected to all other nodes. Typically, a network estimation method is applied to remove non-relevant relations resulting in a sparser network (i.e., each node is no longer connected to all other nodes). The level of sparsity of the GGM can be set using different methods, but the most common method is to apply a variant of the *least absolute shrinkage and selection operator* (LASSO; Tibshirani, 1996) termed *graphical LASSO* (GLASSO; Friedman, Hastie, & Tibshirani, 2008). The GLASSO is a regularization technique that estimates both the model structure and the parameters of a sparse GGM (Epskamp & Fried, 2018). It has a tuning parameter (λ) that is chosen by minimizing the extended Bayesian information criterion (EBIC; Chen & Chen, 2008), which has been shown to accurately retrieve the simulated network structure in Monte Carlo studies (Epskamp & Fried, 2018; Foygel & Drton, 2010).

A newer version of EGA (Golino et al., 2020) uses a different network estimation method called *triangulated maximally filtered graph* method (TMFG; Massara et al., 2016; Christensen, Cotter, & Silvia, 2019). The TMFG method computes a zero-order correlation matrix from the data and applies a constraint on the number of edges to retain in the network ($3n - 6$, where n is the number of nodes; Massara et al., 2016). The algorithm begins by identifying the four variables with largest sum of correlations to all other variables and connects them to each other. Then, variables are iteratively added to the network based on the largest sum of three correlations to nodes already in the network. This process is repeated until all nodes are connected in the network. The resulting network is composed of 3- and 4-node cliques (i.e., sets of connected nodes; a triangle and tetrahedron, respectively). Hereafter, EGA will refer to EGA with GLASSO network estimation and EGAtmfg will refer to EGA with TMFG network.

Once the network is estimated using one of these two methods, the Walktrap community detection algorithm is applied. A detailed introduction of the walktrap algorithm can be found in Golino et al. (2020). In the next paragraphs we summarize how this algorithm works. The Walktrap algorithm uses a process known as *random walks* or stochastic transitions from one node to another over an edge. To define the random walk, let \mathbf{A} be a square matrix of edge weights (e.g., partial correlations) in the network, where \mathbf{A}_{ij} is the strength of the (partial) correlation between node i and j and a node's strength is the sum of node i 's connections to its neighbors $NS = \sum_j A_{ij}$. The steps move from one node to another randomly and uniformly using a transition probability, $\mathbf{P}_{ij} = \frac{\mathbf{A}_{ij}}{NS(i)}$, which forms the transition matrix, \mathbf{P} . The transition matrix is used to compute a distance metric, r , measuring the structural similarity between nodes (Equation 7), and is computed to determine the communities that the nodes belong to. This structural similarity is defined as (Pons & Latapy, 2005):

$$r_{ij} = \sqrt{\sum_{k=1}^n \frac{(\mathbf{P}_{ik} - \mathbf{P}_{jk})^2}{NS(k)}}. \quad (7)$$

This distance can be generalized to the distance between nodes and communities by beginning the random walk at a random node in a community, C . This can be defined as:

$$\mathbf{P}_{C_j} = \frac{1}{|C|} \sum_{i \in C} \mathbf{P}_{ij}. \quad (8)$$

Finally, this can be further generalized to the distance between two communities:

$$r_{C_1 C_2} = \sqrt{\sum_{k=1}^n \frac{(\mathbf{P}_{C_1 k} - \mathbf{P}_{C_2 k})^2}{NS(k)}}, \quad (9)$$

where this definition is consistent with the distance between nodes in the network (Equation 7).

As pointed by Golino et al. (2020), the algorithm begins by having each node as a cluster (i.e., n clusters) and then begins to iteratively choose two clusters after computing the distances, r , between all adjacent nodes. These two clusters chosen are then merged into a new cluster, updating the distances between the node(s) and cluster(s) with each merge (in each $k = n - 1$ steps).

If clusters are adjacent to one another (i.e., an edge between them), then they are merged using an approach based on Ward's agglomerative clustering (Ward, 1963) that depends on the estimation of the squared distances between each node and its community (σ_k), for each k steps of the algorithm. Since computing σ_k is computationally expensive, Pons and Latapy (2005) adopted an efficient approximation that only depends on the nodes and the communities rather than the k steps. The approximation seeks to minimize the variation of σ that would be induced if two clusters (C_1 and C_2) are merged into a new cluster (C_3):

$$\Delta\sigma(C_1, C_2) = \frac{1}{n} \left(\sum_{i \in C_3} r_{iC_3}^2 - \sum_{i \in C_1} r_{iC_1}^2 - \sum_{i \in C_2} r_{iC_2}^2 \right). \quad (10)$$

The resulting values of Ward's approximation adopted by Pons and Latapy (2005) can be stored in a balanced tree, and a sequence of P_k partitions into clusters ($1 \leq k \leq n$, being n the total number of nodes) is obtained. The best number of clusters is defined as the partition that maximizes modularity.

As Golino et al. (2020) note, modularity is a measure that was proposed by Newman (2004) to identify meaningful clusters in networks and can be calculated as follows. Let j and k be two clusters in a network with m and n nodes. If the number of edges between clusters is p , then one-half of fraction of the edges linking j and k is $e_{jk} = \frac{1}{2}p$, so that the total fraction of edges between the two clusters is $e_{jk} + e_{kj}$ (Newman, 2004). Conversely, e_{jj} represents the fraction of edges that fall within cluster j , whose sum

equals one: $\sum_j e_{jj} = 1$. Newman (2004) points out that a division of networks into clusters is meaningful if the value of the sums of e_{jj} and e_{kk} is maximized. However, in cases where only one cluster is presented, the maximal value will be one, which is also the value of $\sum_j e_{jj}$. Therefore, for networks composed by only one cluster this index is not informative. A solution Newman (2004) proposed was to calculate an index that takes $\sum_j e_{jj}$ and subtract from it the value that it would take if edges were placed at random. For a given cluster j , the modularity is calculated as:

$$Q = \sum_j (e_{jj} - a_j^2), \quad (11)$$

where a_j is given by $\sum_k e_{jk}$, which represent the fraction of edges that connect to vertices in cluster j . Therefore, the modularity index penalizes network structures with only one cluster, since in this condition the value of Q would be zero (Newman, 2004).

Golino and Epskamp (2017) studied the accuracy in estimating the number of dimensions of EGA along with six classical techniques in a simulation manipulating the number of factors, number of items, sample size, and correlation between factors. Golino and Epskamp (2017) showed that the most accurate techniques were PA (89% of accurate estimates) and EGA (93%). More recently, Golino et al. (2020) compared EGA and EGAtmfg along with five factor-analytic techniques, including two automated versions of Cattell’s scree test (Cattell, 1966): scree test optimal coordinate (OC) and acceleration factor (AF) methods (Raiche, Walls, Magis, Riopel, & Blais, 2013). The results presented by Golino et al. (2020) showed that EGA (87.91%), parallel analysis (83.01%) and EGAtmfg (74.61%) presented high to moderately high accurate estimations (i.e. correctly estimating the number of simulated factors). Golino et al. (2020) also showed that EGA works well for both dichotomous (overall accuracy of 85%) and continuous data (overall accuracy of 91%), although the accuracy in estimating the number of simulated factors was higher for continuous data.

On the one hand, Golino et al. (2020) found that when EGA and EGAtmfg provided the same estimate for 78% of the datasets, they had a nearly perfect accuracy (91.85%). On the other hand, when the two methods disagreed, the accuracy of both EGA (73.3%) and EGAtmfg (12.9%) diminished. In these cases where EGA and EGAtmfg disagree, it’s important to consider potential alternative solutions (with more or fewer dimensions) to those suggested by the methods. In this case, a fit index specifically designed for dimensionality assessment would play a crucial role, motivating the current article.

Another issue of central importance is the number of steps used in the random walk process of the Walktrap algorithm. To date, EGA has been evaluated using the default value of four steps. Given previous simulation evidence, four steps appears to be adequate for the number of variables typically used in psychological constructs, which has been verified in simulations outside of the area (Pons & Latapy, 2005). Indeed, in Pons and Latapy’s (2005) seminal study, three and four steps appeared to be optimal when there are fewer than 100 variables. Selecting the optimal number of steps is important since it may increase the precision of the EGA estimation, leading to more reliable and interpretable factors. Therefore, developing a fit metric that could be used

to compare the dimensionality structure estimated via EGA with different numbers of steps used in the Walktrap algorithm (e.g., from 3 to 10) could improve the precision of the EGA procedure.

Entropy and the Structure of Multivariate Data

The current paper presents a family of three new fit indices, *Entropy Fit Indices*, which are based on an information-theoretic concept of entropy that can be used to check the fit of the dimensionality structures estimated via EGA and other factor-analytic methods. This section introduces the main concepts used to develop the EFIs, starting with the concept of Shannon’s entropy (Shannon, 1948) and later introducing Von Neumann’s entropy (Von Neumann, 1927). The goal of this section is to help the reader understand the conceptual and technical origins of the entropy fit measures proposed in the current article.

Shannon’s Entropy: Estimation and Properties

The study of entropy originated in physics (specifically, thermodynamics) with the work of Boltzmann and others (Watanabe, 1960). Although a specific type of entropy was being used in communication theory as early as 1928 (Watanabe, 1960), a formal mathematical theory of communication based on entropy was later developed by Shannon (1948). The use of entropy has increased significantly over the years and across many different areas such as communication, neuroscience (Paninski, 2003), biology (Tsuruyama, 2018), ecology (Harte & Newman, 2014), physics (Zurek, 2018), chemistry (Ferenci & Kovács, 2014), and other fields (Volkenstein, 2009).

The entropy of a discrete random variable X characterizes the *uncertainty* associated with X or the degree of disorganization of X (Wiener, 1961). In other words, it depicts the average amount of unpredictability that is removed when the outcome of X is known (Yeung, 2008), which can also be considered a measure of *information* in X (entropy and information are negatively related). Entropy depends on the distribution of $p(X)$, and is maximized when all events of X are equiprobable (i.e., $p(x) = \frac{1}{N}$, where N is the number of elements of X ; Shannon, 1948).

The *maximum likelihood* estimator (MLE) or the *plug-in* method (Antos & Kontoyiannis, 2001) is one of the most common methods to estimate the entropy of a random variable. Let X be a discrete random variable that takes on values in a set $\mathcal{X} = (x_1, x_2, \dots, x_n)$, with a probability mass function $p(x) = P(X = x)$. Antos and Kontoyiannis (2001) showed that considering additive functionals with the general form:

$$F \triangleq g \left(\sum_{x \in \mathcal{X}} f(x, p(x)) \right), \quad (12)$$

where f and g are non-negative arbitrary real-valued functions, the *plug-in estimate* for F is $\hat{F} = F(p_n)$, where p_n is the empirical distribution:

$$p_n(x) = \frac{1}{n} \sum_{j=1}^n I_{(X_j=x)}. \quad (13)$$

The MLE of entropy is, thus:

$$\hat{H}_{(p_n)} \equiv - \sum_{x \in \mathcal{X}} p_n(x) \log p_n(x). \quad (14)$$

Or, more directly:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x). \quad (15)$$

where $p(x)$ is the relative frequency of the elements $x \in \mathcal{X}$.

Entropy can also be extended to multiple variables. Let X and Y be discrete random variables with a joint probability of $p(x, y)$. The entropy of the joint distribution is (Shannon, 1948):

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y). \quad (16)$$

In the case where X and Y are independent variables, their joint entropy will always be less than or equal to the sum of their individual entropies:

$$H(X, Y) \leq H(X) + H(Y). \quad (17)$$

If X and Y are not independent variables, presenting a conditional probability distribution for Y given X as $p(y|x)$, then the conditional entropy $H(Y|X)$ is “the average of the entropy of Y , for each value of X , weighted according to the probability of getting that particular X ” (Shannon, 1948, p. 12). Mezard and Montanari (2009) summarizes the idea of conditional entropy as the amount of information we obtain from the value of y if we know x . The MLE estimation of conditional entropy is:

$$H(Y|X) = - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x). \quad (18)$$

The joint entropy of X and Y can be re-written as the entropy of X plus the entropy of Y given X :

$$H(X, Y) = H(X) + H(Y|X). \quad (19)$$

In other words, the uncertainty of the joint event (X, Y) is the uncertainty of X plus the uncertainty of Y when X is known (Shannon, 1948). Hence, the uncertainty of Y never increases by knowledge of X :

$$H(X) + H(Y) \geq H(X, Y) = H(X) + H(Y|X) \therefore H(Y) \geq H(Y|X). \quad (20)$$

Also, if X and Y are independent, then $H(Y) = H(Y|X)$.

Total Correlation

In his seminal paper, Watanabe (1939) created a modified index of entropy termed *total correlation* to measure not only the uncertainty of random variables, but also the strength of their correlation beyond their average interaction (Watanabe, 1960). The total correlation of a set of variables $\mathcal{X} = (x_1, x_2, \dots, x_n)$ is a non-negative index that increases as entropy of individual variables increases or as entropy of the set of variables (the joint entropy of \mathcal{X}) decreases (Watanabe, 1960), and is calculated as follows:

$$C_{tot_{\mathcal{X}}} = \left(\sum_{i=1}^n H(x_i) \right) - H(x_1, x_2, \dots, x_n) \geq 0. \quad (21)$$

The interpretation of total correlation provided by Watanabe (1960) can be illustrated with a very simple example. Suppose the set of variables \mathcal{X} is divided in two subsets: v containing half the elements of \mathcal{X} and ω containing the other half. Observing the variables contained in v and ω separately, the information in each subset is $H(v)$ and $H(\omega)$. The individual information in each subset of variables individually is higher than the information carried by \mathcal{X} if and only if v and ω are independent. In this scenario of independence, total correlation is the *loss of information* or redundancy. If $H(v)$ and $H(\omega)$ are dependent, then the difference between their joint entropy and the sum of their individual entropies is a measure of the correlation between them, reflecting *ignorance before observation*. Watanabe (1960) points out that without any observation, the ignorance about a subset v is $H(v)$, but after observing v the ignorance becomes $H(\mathcal{X}) - H(v)$. The decrease of ignorance is thus the information about ω provided by the observation of v .

Total correlation was used as the basis for a technique termed *interdependence analysis* (Watanabe, 1969), which replaced a metric of distance to this entropic measure of correlation in cluster identification. The general idea of interdependence analysis is to

partition a set into subsets until interdependence is minimized (Watanabe, 2000). Watanabe (2000) showed that the construction of clusters using total correlation could lead to unwanted problems in the partitioning of a multidimensional space of variables. In particular, by not considering the size of the subsets of a multidimensional space, small subsets could be removed due to their small entropy value, irrespective of the relationship to the other subsets. Watanabe (2000, 2001) introduced a slight modification to the interdependence analysis process by modifying the total correlation index (Equation 21) and replacing it with the *K-function*. Considering a set of variables $\mathcal{X} = (x_1, x_2, \dots, x_n)$, X_v , and X_ω with cardinalities n_1 and n_2 (respectively), the *K-function* representing the interdependence between X_v and X_ω is given by:

$$k(X_v, X_\omega) = n_1 H(X_v) + n_2 H(X_\omega) - (n_1 + n_2) H(X_v, X_\omega). \quad (22)$$

Contrary to the entropy index that is infra-additive (see Equation 17), the product of entropy by the cardinalities is supra-additive, making the *K-function* non-positive:

$$n_1 H(X_v) + n_2 H(X_\omega) \leq (n_1 + n_2) H(X_v, X_\omega) \therefore k(X_v, X_\omega) \leq 0. \quad (23)$$

Using the notation above, the *K-function* is the difference between the average entropy of X_v and X_ω from the entropy of the super-set \mathcal{X} . The *K-function* can be used as a measure of entropy reduction by partitioning through which the minimization is the reduction of the uncertainty or unstableness (Watanabe, 2000, 2001) in the partition of a multidimensional space. It can also be interpreted as the maximization of information gain by the partitioning of a multidimensional space (Watanabe, 2000). Despite its interesting characteristics, the *K-function* could not be used as a fit index for dimensionality assessment (or for clustering) because it decreases with the increase in the number of subgroups of variables (i.e., clusters). Therefore, the application of *K-function* to decide on the number of clusters, factors, or partitionings of a multidimensional space is limited.

The Entropy Fit Indices: Properties and Estimation

Given the limitation of the *K-function* proposed by Watanabe (2000), it is necessary to develop a new metric that does not decrease with the increase in the number of partitions of a multidimensional space (or clusters, factors, and so on). The *Entropy Fit Index (EFI)* was developed as an alternative to the *K-function* with the goal to identify the best structure of a given set of random variables from two or more alternative structures. In other words, EFI can be used to identify the best partitioning of a multidimensional space that better reflects the underlying latent factors without necessarily decreasing with the increase in the number of factors.

Let η be a set with two latent factors ($\eta = (\eta_a, \eta_b)$) with variance-covariance matrix Σ :

$$\Sigma = \Lambda\Phi\Lambda' + \Theta, \quad (24)$$

where *lambda* (Λ) is a $k \times 2$ factor loading matrix for k variables and two factors, *phi* (Φ) is the structure matrix of the latent variables (i.e., a 2×2 matrix of correlations among factors), and *theta* (Θ) is the covariance matrix of the residuals. Now, let η_a and η_b be two latent factors associated with the following matrices of observable scores:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & a_{p3} & \cdots & a_{pn} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1m} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2m} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & b_{p3} & \cdots & b_{pm} \end{bmatrix}$$

where p is the number of participants (subjects), n is the number of variables associated with η_a , and m is the number of variables associated with η_b . Sum scores can be computed as:

$$S_{\eta_a} = \begin{bmatrix} s_{a_1} \\ s_{a_2} \\ s_{a_3} \\ \vdots \\ s_{a_p} \end{bmatrix} \quad (25) \quad S_{\eta_b} = \begin{bmatrix} s_{b_1} \\ s_{b_2} \\ s_{b_3} \\ \vdots \\ s_{b_p} \end{bmatrix} \quad (26)$$

where s_{a_1} and s_{b_1} are the sum scores of variables in \mathbf{A} and \mathbf{B} (respectively) for participant one.

The entropy fit index of the structure formed by η_a and η_b can now be computed as:

$$\text{EFI} = \left[\frac{H(\mathbf{S}_{\eta_a}) + H(S_{\eta_b})}{2} - H(S_{\eta_a}, S_{\eta_b}) \right] + \left[\left(H_{max} - \frac{H(S_{\eta_a}) + H(S_{\eta_b})}{2} \right) \times \sqrt{2} \right], \quad (27)$$

where $H(S_{\eta_a})$ and $H(S_{\eta_b})$ are the entropy of the sum scores of the variables from factors η_a and η_b (respectively), $H(S_{\eta_a}, S_{\eta_b})$ is the joint entropy of the sum scores, and 2 is the number of factors. The entropy of the sum scores for all variables from η_a and η_a is H_{max} . So, H_{max} can be considered as the entropy estimated for a single factor containing all the $n + m$ variables in η_a and η_b . Finally, the difference between the maximum entropy (i.e., entropy estimated for a single factor containing all the $n + m$ variables) and the average entropy of the factors is multiplied by the square root of the number of factors (from now on this difference will be termed *difference entropy*).

In sum, Equation 27 calculates the difference between the average entropy for the individual factors and the joint entropy of the factors, while including a penalization

for the number of factors. The penalization modulates the EFI to not decrease with an increase in the number of factors, except if each additional factor decreases the uncertainty or disorder of the system of variables. Equation 27 can be easily extended to N_F factors:

$$\text{EFI} = \left[\frac{\sum_{i=1}^{N_F} H(S_{\eta_i})}{N_F} - H(S_{\eta_1}, S_{\eta_2}, \dots, S_{\eta_i}) \right] + \left[(H_{max} - \frac{\sum_{i=1}^{N_F} H(S_{\eta_i})}{N_F}) \times \sqrt{N_F} \right] \quad (28)$$

The EFI can be used as a measure of entropy reduction by partitioning of a multidimensional space into several groups of variables, in which the minimization is the minimization of the uncertainty or disorder due to the correct identification of the underlying factors. In contrast with the *K-function*, EFI does not decrease with the increase in the number of factors. The EFI will be lower if the correct structure is provided (i.e., the correct number of factors and the correct composition of items per factor), irrespective of the number of factors. This principle is due to the *entropy and the fineness of classification* (e.g., classifying items into subgroups) theorem (Watanabe, 2000) that states the following:

Theorem 1 (*Entropy and the Fineness of Classification*). *Let X_I and X_J be subsets of X , and A be a fixed set of predicates. If a classification of A with respect to the latter is a refinement of the classification with respect to the former then the formal entropy of X_I is never larger than that of X_J .*

Theorem 1 implies that a structure that splits a set of random variables into groups, which reflect the underlying latent factors, will have a lower entropy than a structure that does not reflect the latent factors. An example helps illustrate this principle. Let η_1 and η_2 be two latent factors with four variables each. For every unique combination of the items into two sets of four, the EFI is recorded, generating the result showed in Figure 1. The first and last elements of the x-axis represent the correct structure (1, 1, 1, 1, 2, 2, 2, 2 or 2, 2, 2, 2, 1, 1, 1, 1), while the other elements of the x-axis presents all the other possible combinations of the variables into two factors (e.g., 1, 1, 1, 2, 1, 2, 2, 2). At the bottom of Figure 1, a matrix representing the distribution of the variables into two factors (black representing variables from factor η_1 , white representing variables from factor η_2) is shown. The columns of the matrix show the distribution of the items into two groups, while the rows are divided into two parts, the first four rows represents η_1 , the last four rows represents η_2 .

Figure 1 shows a perfect symmetric relationship between the distribution of variables per factor and the EFI index (y-axis). The EFI is the same for the structure (1, 1, 1, 1, 2, 2, 2, 2) and (2, 2, 2, 2, 1, 1, 1, 1). Also, (1, 1, 1, 2, 1, 2, 2, 2), second element in the x-axis, and (2, 2, 2, 1, 2, 1, 1, 1), 69th element of the x-axis, has the same EFI. This symmetric behavior of EFI can be explained by the nature of the entropy measures based on total correlation or on the *K-function*: it reflects the relationship among groups of variables rather than the relationship between variables. Thus, every time a group of variables is classified in an incorrect way (i.e., not reflecting the true underlying latent factors) there is loss of information and the EFI increases. For example, if EFI were computed in the same data used to generate Figure 1, but with a three-factor structure instead of two (i.e., 1, 1, 1, 1, 2, 2, 3, 3), the EFI would increase from -0.402 to -0.328.

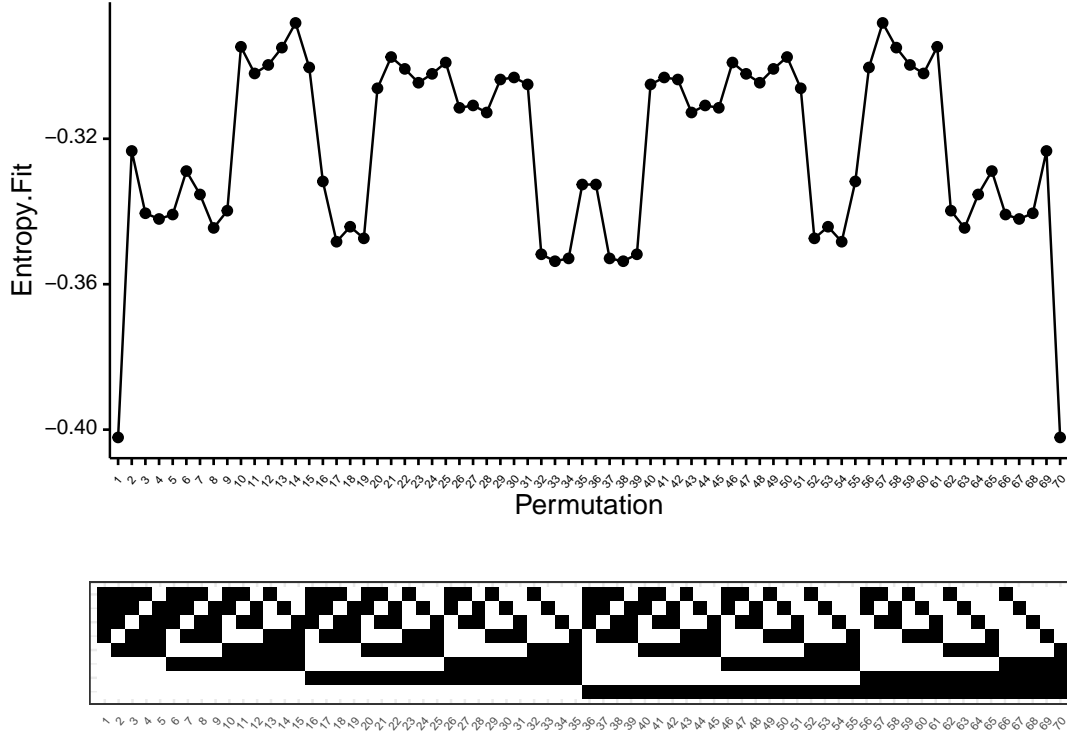


Figure 1. Entropy fit index per combination of variables into two factors

It is important to note that Figure 1 focuses on a multidimensional structure with two factors and all combinations of variables in two groups of four variables. The pattern is different if we vary the number of factors, in which the EFI value would be lower only if the correct structure is provided, increasing linearly with the number of factors. This is an important difference between EFI and the *K-function* since the latter always decreases with the increase in the number of groups of variables (or factors).

To illustrate this principle, consider a multidimensional structure with 100 factors, four variables per factor, and a process with 100 steps. In the first steps, the EFI, total correlation, *K-function*, and the difference between H_{max} and the average entropy of the factors (*difference entropy*) is recorded using the correct structure (i.e., 100 factors, four variables per factor). In each subsequent step one factor is split into two sub-factors, with two variables each and all the indices used in the first step are recorded. The result of this process shows the variability of the EFI, total correlation, *K-function*, and difference entropy as the factors are sequentially sub-divided into two factors.

Figure 2 shows exactly the pattern described above. The *K-Function* of the factors (or average entropy of the factors) decreasing with the number of latent variables, while the difference entropy increases following an exponential function (see Equation 29).

$$H_{Dif} \simeq 1.1568 \times \frac{e^{(0.0335 \times N_F)}}{1 + e^{(0.0335 \times N_F)}}. \quad (29)$$

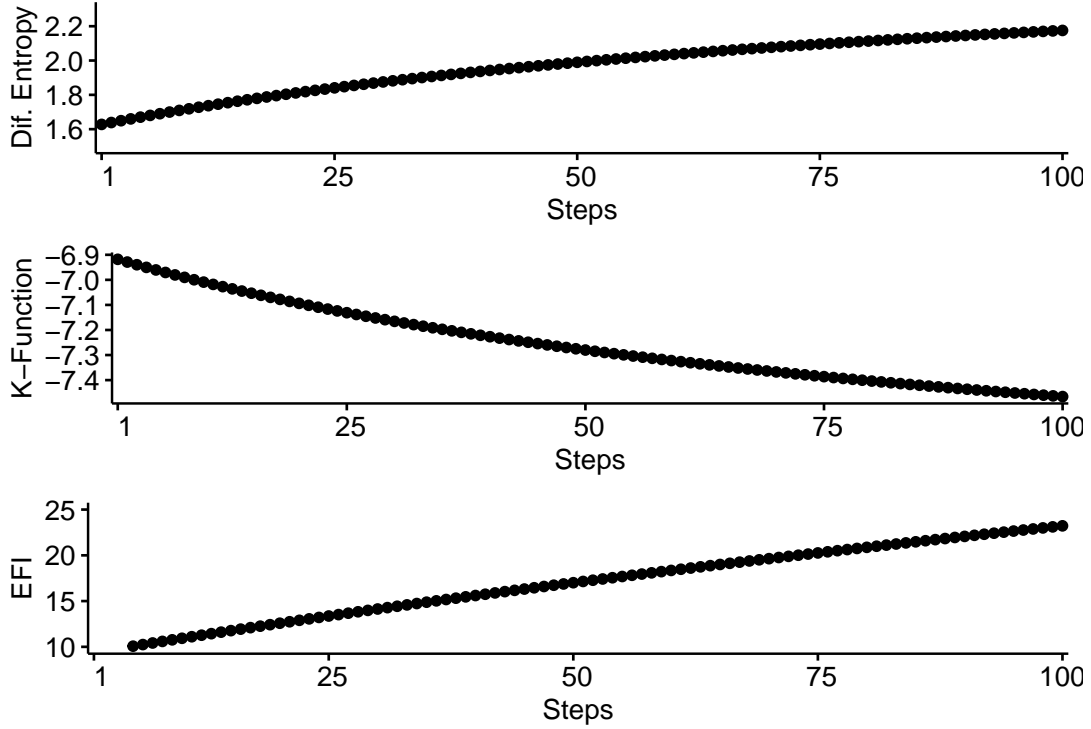


Figure 2. Entropy fit index per combination of variables into two factors

In sum, what Figure 2 shows is that how EFI differs from the K -function and makes it more suitable as a measure of fit due to the difference entropy, which penalizes the number of partitions of the multidimensional space or number of factors.

In terms of estimation, several estimators are available for entropy (especially empirical entropy or Shannon entropy; Paninski, 2003). In the current article, the maximum likelihood estimator of entropy (Equation 15) is used to calculate the EFI. In the MLE of entropy, the number of bins to compute the empirical distribution of the continuous variables needs to be provided. The optimal number of bins can be estimated using the formula suggested by Cellucci, Albano, and Rapp (2005), where the number of bins is the greatest integer such that:

$$N_{Bins} \leq \sqrt{\frac{N}{5}}, \quad (30)$$

where N is the sample size. So, after calculating the number of bins, the sum score of each group of variables (indicating a specific latent factor) is partitioned into N_{Bins} elements, and Equation 28 is computed. It is important to note that the *plug-in* estimator of entropy, as used in the current paper, can only be computed with raw data.

Entropy Fit Index with Von Neumann’s Entropy and the Total Entropy Index: Estimations using a correlation matrix

Because it’s common in the areas of psychology, education, health and related fields for a researcher to have access only to the correlation matrix of multivariate datasets, it is valuable to have an alternative estimation for the EFI that could be computed using correlations. A plausible and approximate solution is to use matrix algebra to calculate Von Neumann entropy (Von Neumann, 1927)—an index that was developed to quantify the amount of disorder in a system. It’s also been used to quantify the entanglement between two subsystems in quantum physics (Preskill, 2018), which occurs when two (or more) particles become inextricably linked. The entanglement of two photons can be verified using specific visualizations of the spatial correlations and momentum correlations (Moreau et al., 2019). In this case, the visualization provides a clue to how linked (entangled or mixed) two systems (or particles) are. The entanglement of two or more systems can also be expressed in terms of a *density matrix*, a type of matrix used in quantum mechanics that is akin to the probability distribution of position and momentum (i.e., phase-state probability) in classical statistical mechanics (Hall, 2013). Any density matrix has three main characteristics: (1) it is real symmetric, (2) positive semi-definite, and (3) has trace equal to one.

Preskill (2018) shows that for any density matrix ρ , Von Neumann entropy is:

$$\mathcal{S}(\rho) = -tr(\rho \log \rho). \tag{31}$$

Any correlation matrix can be transformed into a density-like matrix by using the number of variables as a scaling factor that will make the trace of the matrix equal to one (Anderson, 1963). By dividing the elements of the matrix by the number of variables, the correlation matrix will hold the same properties as any density matrix.

Wihler, Bessire, and Stefanov (2014) shows that eigenvalues can be used to estimate Von Neumann entropy. Given a density matrix ρ with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m \geq 0$, Von Neumann entropy is:

$$\mathcal{S}(\rho) = -tr(\mathcal{L}(\mathbf{D})), \tag{32}$$

where \mathcal{L} is a continuous function on the real interval $[0, \infty)$ that takes values $x \log(x)$ if $x \geq 0$ and 0 if $x = 0$, and $\mathcal{L}(\mathbf{D}) = diag(\mathcal{L}(\lambda_1), \mathcal{L}(\lambda_2), \dots, \mathcal{L}(\lambda_m))$. In other words, the Von Neumann entropy of a density matrix is the Shannon entropy of the vector of eigenvalues (Preskill, 2018).

As presented earlier, to calculate the entropy fit index of a given structure using Shannon’s entropy the joint entropy of the factors must be computed. The joint Von Neumann entropy of two (or more) density matrices can be calculated as the Kronecker product of the matrices. Given two density matrices, ρ_A and ρ_B , the product state ρ_{AB} is given by $\rho_A \otimes \rho_b$ (Preskill, 2018). Therefore, after calculating ρ_{AB} , Equation 32 can be used to estimate the Von Neumann entropy of the joint state (or the joint Von Neumann entropy). If the number of joint states is large (e.g., more than five), the resulting Kronecker product can be very large. To avoid this issue, the product

of the individual eigenvalues can be used instead of the Kronecker products, since the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ is $\lambda_1\mu_1, \dots, \lambda_1\mu_m, \dots, \lambda_2\mu_1, \dots, \lambda_2\mu_m, \dots, \lambda_n\mu_m$ (Laub, 2005).

Since the EFI is estimated using the raw data (i.e., Shannon's entropy), while Von Neumann entropy is estimated using a density-like, scaled version of a correlation matrix, two modifications from Equation 28 are necessary to estimate the EFI Von Neumann (EFI.vn; see Equation 33). First, the sum scores of variables belonging to the factors are replaced by sub-matrices of $\boldsymbol{\rho}$. So, two factors (η_a and η_b) can be represented as a different combination of sub-elements (rows and columns) from $\boldsymbol{\rho}$: as $\boldsymbol{\rho}_a$ and $\boldsymbol{\rho}_b$. If the first factor is composed by variables i, j and k of $\boldsymbol{\rho}$, the sub-matrix $\boldsymbol{\rho}_a$ will be:

$$\boldsymbol{\rho}_a = \begin{bmatrix} a_{ii} & a_{ij} & a_{ik} \\ a_{ji} & a_{jj} & a_{jk} \\ a_{ki} & a_{kj} & a_{kk} \end{bmatrix}$$

The second factor being composed by variables m, n and o of $\boldsymbol{\rho}$, the sub-matrix $\boldsymbol{\rho}_b$ will be:

$$\boldsymbol{\rho}_b = \begin{bmatrix} b_{mm} & b_{mn} & b_{mo} \\ b_{nm} & b_{nn} & b_{no} \\ b_{om} & b_{on} & b_{oo} \end{bmatrix}$$

The Von Neumann entropy of $\boldsymbol{\rho}_a$ and $\boldsymbol{\rho}_b$ are $\mathcal{S}(\boldsymbol{\rho}_a)$ and $\mathcal{S}(\boldsymbol{\rho}_b)$, respectively.

The second modification relates to the penalization in which the maximum entropy is replaced by the total entropy of the matrix $\mathcal{S}(\boldsymbol{\rho})$. Also, the penalization now is subtracted instead of added as in Equation 28. EFI.vn can be computed for two factors as follows:

$$EFI_{VN} = \left[\frac{\mathcal{S}(\boldsymbol{\rho}_a) + \mathcal{S}(\boldsymbol{\rho}_b)}{2} - \mathcal{S}(\boldsymbol{\rho}_a \otimes \boldsymbol{\rho}_b) \right] - \left[\left(\mathcal{S}(\boldsymbol{\rho}) - \frac{\mathcal{S}(\boldsymbol{\rho}_a) + \mathcal{S}(\boldsymbol{\rho}_b)}{2} \right) \times \sqrt{2} \right]. \quad (33)$$

where $\mathcal{S}(\boldsymbol{\rho}_a)$ and $\mathcal{S}(\boldsymbol{\rho}_b)$ are the Von Neumann entropies for factors η_a and η_b , $\mathcal{S}(\boldsymbol{\rho})$ is the total entropy of the matrix $\boldsymbol{\rho}$ and $\mathcal{S}(\boldsymbol{\rho}_a \otimes \boldsymbol{\rho}_b)$ is the joint entropy of the factors, which can be calculated using the Kronecker product of the matrices or (in case of multiple factors) as the product of eigenvalues, as pointed out earlier. Equation 33 can be easily expanded to N_F factors:

$$EFI_{VN} = \left[\frac{\sum_{i=1}^{N_F} \mathcal{S}(\boldsymbol{\rho}_i)}{N_F} - \mathcal{S}(\boldsymbol{\rho}_1 \otimes \boldsymbol{\rho}_2 \otimes \dots \otimes \boldsymbol{\rho}_{N_F}) \right] - \left[\left(\mathcal{S}(\boldsymbol{\rho}) - \frac{\sum_{i=1}^{N_F} \mathcal{S}(\boldsymbol{\rho}_i)}{N_F} \right) \times \sqrt{N_F} \right] \quad (34)$$

The use of the joint entropy of the factors can be a possible limitation for EFI.vn. In the EFI the joint entropy is estimated directly using the raw data, but in the EFI.vn it's calculated using the Kronecker product of the individual factors (or the product of the individual eigenvalues per factor, as shown above). This strategy seems useful to detect superposition of states (in the present case the superposition is when two factors are merged into one) since this is a typical property of quantum information theory (Preskill, 2018; Zhao, Yang, Hung, & Li, 2015). However, if one pure state (one factor) is split into two, generating a plausible but incorrect structure, then the performance of the EFI.vn may not be very accurate.

An alternative approach to EFI.vn is to use the total entropy in the correlation matrix, which can be calculated using Equation 32, instead of the joint entropy of the factors in the left side of Equation 33. This alternative EFI.vn measure, hereafter named *total entropy fit index with Von Neumann entropy* (TEFI.vn)¹, maintains the same properties as EFI and EFI.vn by modifying the penalization using only the sum of the factor entropies instead of the mean. Then, TEFI.vn can be calculated as:

$$TEFI_{VN} = \left[\frac{\sum_{i=1}^{N_F} \mathcal{S}(\rho_i)}{N_F} - \mathcal{S}(\rho) \right] + \left[\left(\mathcal{S}(\rho) - \sum_{i=1}^{N_F} \mathcal{S}(\rho_i) \right) \times \sqrt{N_F} \right]. \quad (35)$$

In sum, the entropy fit indices (EFI, EFI.vn and TEFI.vn) represent a combination of both a theoretic framework and empirical refinement. For example, Equation 27 (and the general case in equation 28) can be broken into two components, separated as $EFI = [A] + [B]$. Component $[A]$ is similar to that of the mutual information of a two variables $I(A, B) = H(A) + H(B) - H(A, B)$ (or the total correlation of multiple variables), with a modification that divide the summation on individual entropies by the number of factors yielding what Watanabe (2001) termed 'K-function': $K_{(A,B)} = [(H(A)+H(B))/2] - H(A, B)$. This is the theoretical basis for this section. However, in development we noticed that the K-function alone was not enough to separate factors as $K_{(A,B)}$ strongly favors having a larger number of dimensions. This is because the entropy within individual factors showed a decrease as the number of items included in the sum scores within a factor decreased. That is $H(A)$ and $H(B)$ tend to be lower if less items are used to calculate the distribution of sum scores defining A and B . Meaning that this section of $K_{(A,B)}$ is inflated with the number of factors given a constant number of items.

To account for this, we included a component $[B]$ that reduces the influence of $[A]$ by the number of factors used to describe a given data set. That is, while $[A]$ is expected to decrease monotonically as the number of factors increases, $[B]$ is expected to increase as the number of factors increase. $[B]$ is a measure of the difference from the theoretically maximally entropic factor structure (a single factor) versus the average amount of entropy seen across individual factors of a given factor structure. In other words, $[B]$ represents the reduction in average entropy of a set of data conditional on a given factor structure. The square root of the number of factors was chosen in $[B]$

¹It is important to note that in the TEFI.vn code used in the current paper, entropy is estimated as the negative of the trace of the product of the density matrix by the log of elements of the density matrix, instead of using the matrix logarithm or the eigenvalues of the density matrix (as implemented to calculate EFI.vn). Therefore, in the current implementation, TEFI.vn is based on an entropy-like quantity. We should explore this difference in future papers.

in order to control the expected growth trajectory of $[B]$ as the number of factors increases. We would expect that the effect of adding an additional factor would be conditional on the number of factors already being estimated in the model, showing a decreasing effect as the number of factors increases. That is, the expected decrease in total entropy going from 1 to 2 factors would be higher than the expected decrease in entropy going from 100 to 101 factors. Multiplication by the square root of the number of factors models this behavior.

Methods

Two Monte-Carlo simulations were used to compare the entropy fit indices (i.e., *EFI*, *EFI.vn* and *TEFI.vn*) with the traditional fit indices used in factor models (CFI, RMSEA, TLI, and SRMR). The goal of the first simulation was to check which fit index is more accurate in detecting the correct structure simulated, assuming the number of latent factors is known. In other words, given two latent factors with four variables each, which fit index is the most accurate in identifying the correct placement of variables within the two factors?

The second Monte-Carlo simulation, on the other hand, aims to compare the entropy fit indices and the traditional fit indices in multiple scenarios, varying, for example, the number of factors (same, less, or more than the population structure used to generate the observable data) and how the items are placed in each factor (randomly or not). This second simulation was implemented to answer the following question: what happens with the fit indices (in terms of their accuracy to correctly identify the number of factors and their item composition) when the dimensionality structure being investigated contains the same number of factors, less factors, or more factors than the population structure? Although the first simulation emulates conditions that are not commonly observed in empirical studies (we never actually know the number of factors), the second simulation presents a more realistic set of conditions.

Design

To investigate the adequacy of the entropy fit indices (i.e., *EFI*, *EFI.vn* and *TEFI.vn*) and the traditional fit indices used in factor models (CFI, RMSEA, TLI, and SRMR) two Monte-Carlo simulations are presented. The first simulation identifies the best fit indices under the assumption that the correct number of factors is *known*. In other words, given two latent factors with four variables each, which fit index is the most accurate in identifying the correct placement of variables within the two factors? To answer this question, a permutation strategy is implemented, similar to the first example showed in the last section. The fit indices are computed for every unique combination of the items into two sets of four (in a total of 70 unique combinations, from (1,1,1,1,2,2,2,2), (1,1,1,2,1,2,2,2), . . . , to (2,2,2,1,2,1,1,1), (2,2,2,2,1,1,1,1). Four between-subject data factors were systematically manipulated: the skew of the items (-2; 2; -2 and 2; -2 to 2 in increments of 0.5), sample size (1000 and 5000), factor loadings (.40, .55 and .70), and factor correlations (0, .5 and .7).

In the second Monte-Carlo simulation, the fit indices will be compared in terms of their capacity to identify the correct structure (number of factors and the correct placement of items per factor) in four different set of conditions (shuffled, underfactoring,

overfactoring with mixed items, and overfactoring with a more extreme condition; i.e., no mixed items). For each set of conditions, the fit index for the correct structure (i.e., the structure that corresponds exactly to the simulated population structure; see Figure 3.A and Figure 4.A) is compared to an incorrect structure. The incorrect structures were set as follows: One structure presents the same number of factors as the population structure, but with a random distribution or placement of items per factor (*shuffled condition*; Figure 3.B and Figure 4.B). A second structure reflects the number of factors in the population minus one (*underfactoring condition*; Figure 3.C and Figure 4.C), while a third organization presents the same underfactoring structure, but with a random assignment of items per factor (*underfactoring condition with random placement of items*; Figure 3.D and Figure 4.D).

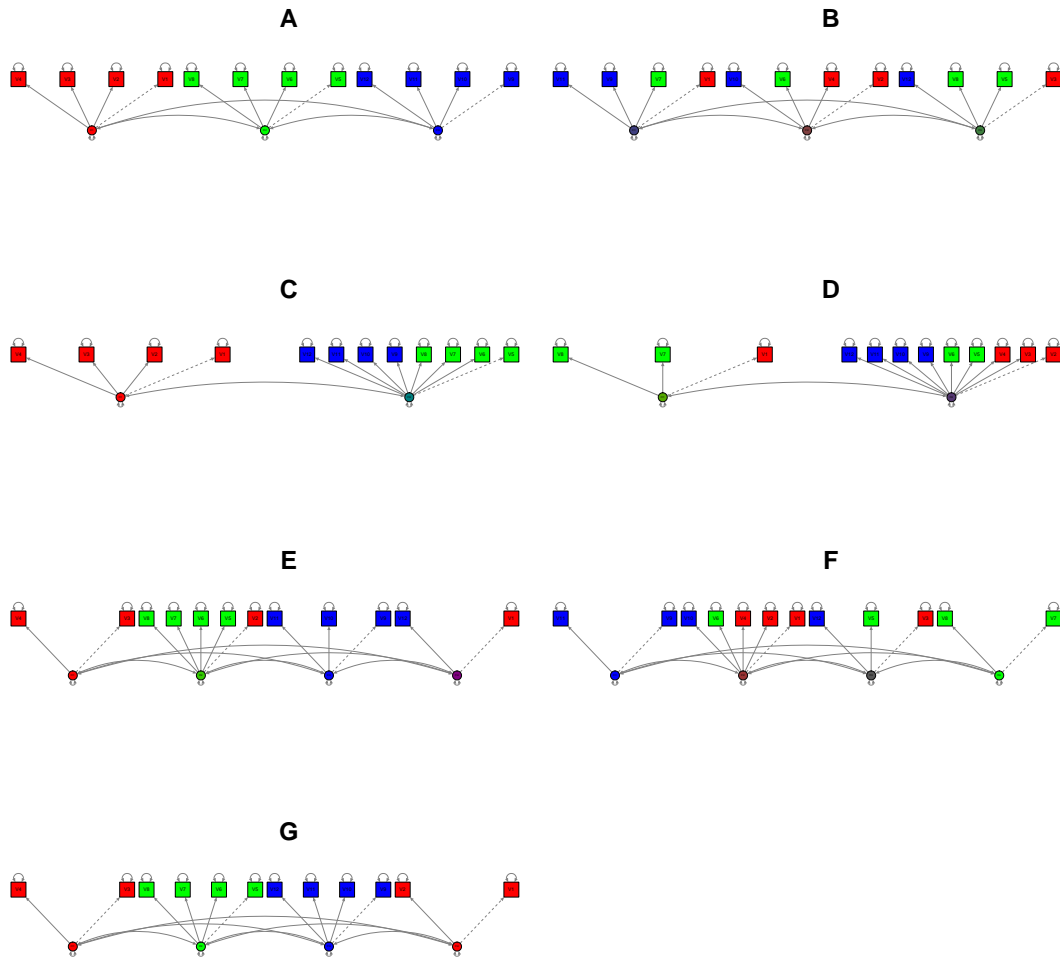


Figure 3. Population structure and examples of models tested, 4 items per factor

The fifth and sixth structures reflect the number of factors in the population plus one (*overfactoring condition with mixed items*; Figure 3.E and Figure 4.E), with only one pure factor (containing items from one single factor) and a random assignment of items per factor (*overfactoring condition with random placement of items*; Figure 3.F

and Figure 4.F). The last condition reflects the number of factors in the population plus one with a more extreme condition where the factors contain items from one single dimension (*overfactoring condition with a more extreme condition*; i.e., no mixed items; Figure 3.G and Figure 4.G).

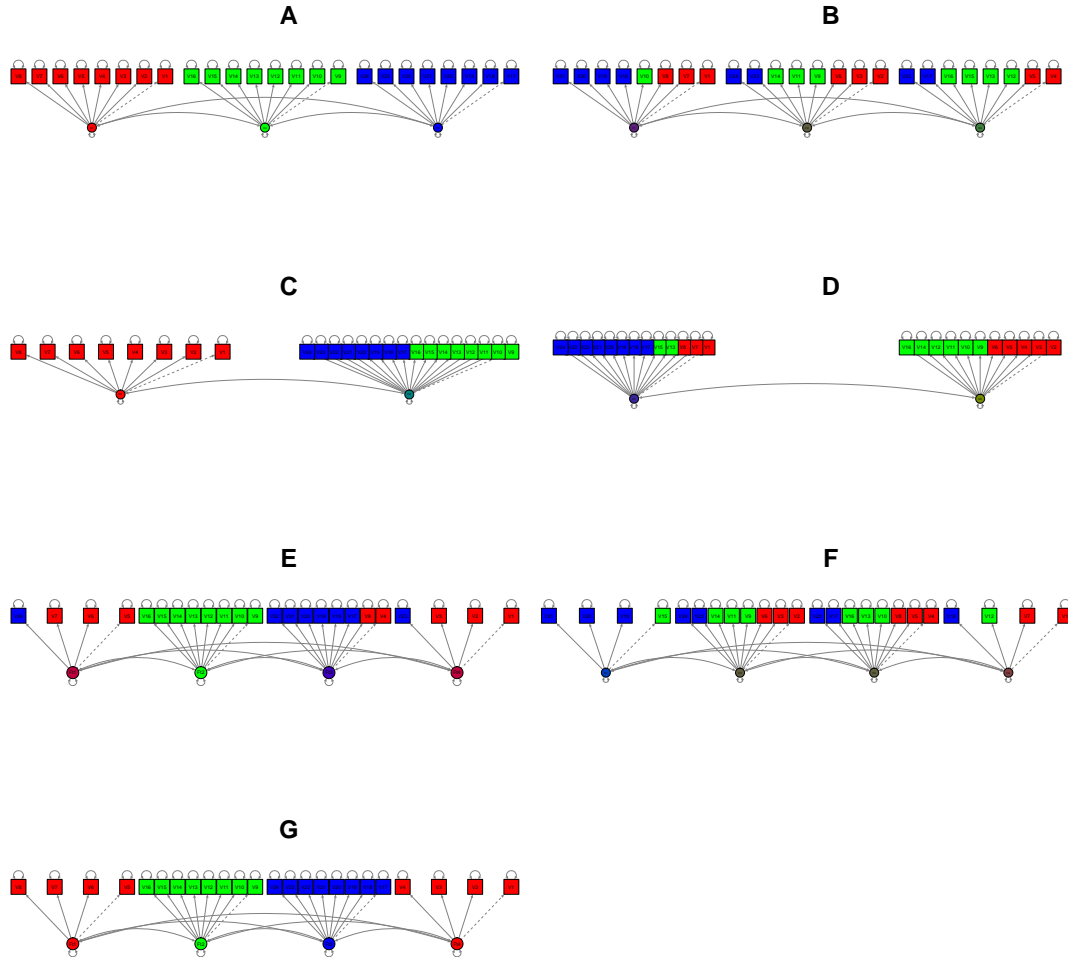


Figure 4. Population structure and examples of models tested, 8 items per factor

Finally, four between-subject data factors were systematically manipulated in the second simulation: factor loadings (.40, .55, and .70), factor correlations (0, 0.5, and 0.7), number of items per factor (4 and 8) and sample size (1000 and 5000). The number of factors was held constant (i.e., three factors).

Factor loadings of .40, .55, and .70 can be considered as poor, good, and excellent, respectively; thus, representing a wide range of factor saturations (Comrey & Lee, 2016). The factor correlations simulated include the orthogonal (.00), large (.50), and very large (.70) factor correlations, according to Cohen (1988). The reason to simulate data with very high factor correlations (.70) is because in some areas within psychology (e.g., intelligence) researchers sometimes have to distinguish between constructs that

are highly correlated (e.g., Kane, Hambrick, & Conway, 2005). The factors generated were composed of four and eight indicators with salient loadings, following previous simulation studies in dimensionality assessment (Golino & Epskamp, 2017; Golino et al., 2020; Velicer, 1976; Widaman, 1993). Sample sizes 1,000 and 5,000 can be considered large (Li, 2016) and very large with the latter allowing for the evaluation of the dimensionality methods in conditions that can approximate their population performance. Additionally, these sample sizes were selected by taking into account that tetrachoric correlations require large sample sizes to achieve acceptable sampling errors (Timmerman & Lorenzo-Seva, 2011).

Data Generation

For each of part of the simulation, 500 sample data matrices of binary variables were generated according to a common factor model procedure that works as follows. First, the reproduced population correlation matrix (with communalities in the diagonal) was computed:

$$\mathbf{R}_R = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}', \quad (36)$$

where \mathbf{R}_R is the reproduced population correlation matrix, $\mathbf{\Lambda}$ is a $k \times r$ factor loading matrix for k variables and r factors, and $\mathbf{\Phi}$ is the structure matrix of the latent variables (i.e., a $r \times r$ matrix of correlations among factors). This procedure implies that the generated data does not contain correlated residuals (minor factors) at the population level.

The population correlation matrix \mathbf{R}_P was then obtained by inserting unities in the diagonal of \mathbf{R}_R , thereby raising the matrix to full rank. The next step was performing a Cholesky decomposition of \mathbf{R}_P , such that:

$$\mathbf{R}_P = \mathbf{U}'\mathbf{U}. \quad (37)$$

If either \mathbf{R}_P was not semi-positive definite (i.e., at least one eigenvalue was ≤ 0) or an item's communality was greater than 0.90, the $\mathbf{\Lambda}$ matrix was replaced and a new \mathbf{R}_P matrix was computed following the same procedure. Subsequently, the sample data matrix of continuous variables was computed as:

$$\mathbf{X} = \mathbf{Z}\mathbf{U}, \quad (38)$$

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.

The resulting continuous variables were dichotomized by applying a set of thresholds according to specific levels of skewness. In the first part of the simulation (the permutation of items per factors), the following levels of skewness were used: -2 , -2

and 2, from -2 to 2 in increments of 0.5 , and 2 . In the second part of the simulation, because binary variables are not necessarily symmetrically distributed in practice, the level of skewness of each item was randomly assigned with equal probability from a set of possible values that ranged from -2 to 2 in steps of 0.50 . The thresholds used to generate these levels of skewness for the binary variables were taken from (Garrido, Abad, & Ponsoda, 2011; Garrido et al., 2013).

It is common in practice to find complex structures in which items present non-zero loadings in multiple factors. To generate cross-loadings with magnitudes consistent to those commonly found in real data (Bollmann, Heene, Küchenhoff, & Bühner, 2015), the procedure described in Meade (2008) was followed for the second part of the simulation: cross-loadings were randomly drawn from a normal distribution (with mean zero and variance of $.05$) for all the items except for the first two in each factor, which were set as markers (i.e., all of their cross-loadings were fixed to zero). Of note, regarding the generation of the main loadings: The function generates the main loadings by drawing random values from a uniform distribution that has a range of ± 0.10 from the specified value. For example, if the main loadings are set at 0.70 , the function generates loading values between 0.60 and 0.80 .

Data Analysis

We used *R* (R Core Team, 2017) for all our analyses. The *EFI*, *EFI.vn*, and *TEFI.vn* were computed using the *EGAnet* package (Golino & Christensen, 2019), while *CFI*, *RMSEA*, *TLI*, and *SRMR* were calculated using *lavaan* (Rosseel, 2012). Confirmatory factor models were estimated using the *WLSMV* estimator. The choice to use the *WLSMV* estimator was due to the discrete nature of the simulated data (dichotomous) since this estimator is theoretically justified for the factor analysis of discrete data and it gives the best results in simulation studies (Barendse et al., 2015; Beauducél & Herzberg, 2006). The figures were generated using the *ggplot2* package (Wickham, 2016) and the *ggpubr* package (Kassambara, 2018). To estimate the *EFI.vn* and *TEFI.vn*, a tetrachoric correlation was computed using the *qgraph* package (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012).

In order to evaluate the performance of the fit indices, the percentage of correct structure selection (**PC**) was used:

$$PC = \frac{\sum C}{N}, \text{ for } C = \left\{ \begin{array}{l} 1 \text{ if } S_{selected} = S_{correct} \\ 0 \text{ if } S_{selected} \neq S_{correct} \end{array} \right\}, \quad (39)$$

where N is the number of sample data matrices simulated, $S_{selected}$ is the selected structure and $S_{correct}$ is the correct structure. The PC criterion has boundaries of 0% and 100% , with 0% signaling complete inaccuracy and 100% indicating perfect accuracy. Since in the current simulation only one structure out of seven is correct, PC has a theoretical base rate near 14% .

The strategies used to select a structure depend on the fit index used. For *EFI* and *EFI.vn*, the selected structure is always the one presenting the lower value. For the other indices, two main strategies were used. The first one is based on the cut-off values only: *CFI* and *TLI* $\geq .95$, *RMSEA* and *SRMR* $\leq .05$. *CFI* and *TLI* values equal to or

greater than .95 can be considered to reflect an excellent fit to the data (Garrido et al., 2016; Hu & Bentler, 1999b; Marsh et al., 2004) as well as RMSEA and SRMR values equal to or lower than .05 (Browne & Cudeck, 1992; Chen et al., 2008; Garrido et al., 2016; Hu & Bentler, 1999b; Marsh et al., 2004). If a structure other than the correct structure presented a value that is considered adequate (passes the cut-off points), then the fit index received a score of zero (i.e., failed to identify the correct structure). The second strategy uses the fit indices as relative measures of fit. Considering all the conditions where the correct structure presented an adequate fit value, the structure with the highest (CFI and TLI) or lowest (RMSEA and SRMR) fit values is selected. If the selected structure is not the correct one, the respective score is zero. Scores of one are only achieved when the selected structure is the correct structure. From now on, when the traditional fit indices are used as relative measures of fit they will be referenced using the abbreviation *.rel*: CFI.rel, TLI.rel, RMSEA.rel and SRMR.rel.

For reproducibility purposes, all R scripts used in the current article, as well as the Rmarkdown file with the manuscript and codes used in the analysis of the result, are available in an online repository at the Open Science Framework platform [here](#).

Results

Simulation 1: Combination of variables into two factors

The results of the first simulation shows that the fit indices can be separated into two groups: one with high PC (TEFI.vn, EFI.vn, SRMR.rel, EFI, TLI.rel, RMSEA.rel and CFI.rel) and the other with low PC (TLI, CFI, RMSEA, SRMR). The most accurate fit indices were (from best to worst): TEFI.vn (94.08%), EFI.vn (93.58%), SRMR.rel (88.80%), EFI (86.02%), TLI.rel (84.96%), RMSEA.rel (83.87%) and (CFI.rel 82.17%). The least accurate fit indices were (from best to worst): TLI (65.51%), CFI (57.45%), RMSEA (32.43%), and SRMR (27.31%). Figure 5 shows the distribution of the percentage correct per fit index.

Figure 6 shows how the accuracy varies per level of factor loadings and correlations. In general, the accuracy improves with the increase in factor loadings (overall accuracy of 54.41% for loadings of .40, 76.29% for loadings of .55 and 86.45% for loadings of .70) and decreases with higher factor correlation (overall accuracy of 87.41% for correlations of 0, 76.96% for correlations of .50 and 52.87% for correlations of .70).

TEFI.vn and EFI.vn presented an almost perfect accuracy for factor loadings of .55 and .70 ($PC_{TEFI.vn} = 98.41\%$, $PC_{EFI.vn} = 98.24\%$ and $PC_{TEFI.vn} = 99.96\%$, $PC_{EFI.vn} = 99.92\%$, respectively), and an impressive accuracy for factor loadings of .40 ($PC_{TEFI.vn} = 83.87\%$, $PC_{EFI.vn} = 82.56\%$). EFI presented a good accuracy for a loading of .40 (74.73%), and high accuracies for loadings of .55 and .70 (89.92% and 93.40%, respectively). TLI.rel, RMSEA.rel and CFI.rel presented lower accuracies for small loadings (68.13%, 62.78% and 59.86%, respectively), but high accuracies for loadings of .55 (93.44%, 92.38% and 90.79% respectively) and .70 (93.32%, 96.46% and 95.86% respectively).

Considering the size of the correlation, it is interesting to note that TEFI.vn, EFI.vn and TLI.rel presented a moderately high accuracy, even when the correlation is .70 ($PC_{TEFI.vn} = 87.61\%$, $PC_{EFI.vn} = 86.97\%$ and $PC_{TLI.rel} = 80.02\%$), followed by SRMR.rel (78.51%). EFI was the fifth best fit measure when correlations between

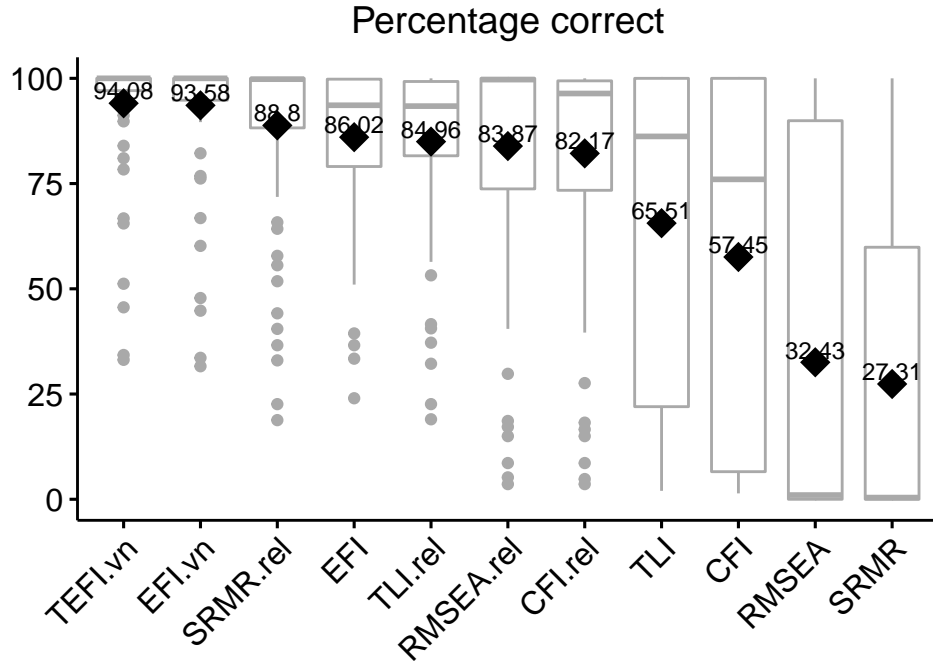


Figure 5. Percentage Correct per Fit Index. The horizontal lines indicate the median values, the numbers inside the plots are the mean values.

factors was .70 (74.43%), followed by RMSEA.rel (70.82%) and CFI.rel (70.78%). textcoloredTable 1 and table 2 shows the mean percentage correct and its 95% confidence interval for each level of correlation and factor loading, respectively.

Table 1.: Mean percentage correct and its confidence interval for each level of correlation.

Method	Correlations	Mean	95% C.I. - Lower	95% C.I. - Upper
TEFI.vn	0	99.25	99.09	99.40
TEFI.vn	50	95.41	95.04	95.79
TEFI.vn	70	87.61	87.02	88.20
EFI.vn	0	98.85	98.66	99.04
EFI.vn	50	94.91	94.51	95.30
EFI.vn	70	86.97	86.36	87.57
SRMR.rel	0	96.76	96.44	97.07
SRMR.rel	50	91.20	90.70	91.71
SRMR.rel	70	78.51	77.78	79.25
EFI	0	96.63	96.31	96.96
EFI	50	86.99	86.39	87.59
EFI	70	74.43	73.65	75.21

RMSEA.rel	0	94.09	93.67	94.52
RMSEA.rel	50	86.78	86.18	87.39
RMSEA.rel	70	70.82	70.01	71.64
TLI.rel	0	86.15	85.53	86.77
TLI.rel	50	88.72	88.16	89.29
TLI.rel	70	80.02	79.30	80.73
CFI.rel	0	89.46	88.91	90.01
CFI.rel	50	86.27	85.65	86.88
CFI.rel	70	70.78	69.97	71.60
TLI	0	94.73	94.33	95.13
TLI	50	81.09	80.39	81.79
TLI	70	20.96	20.23	21.69
CFI	0	93.75	93.32	94.18
CFI	50	74.18	73.40	74.97
CFI	70	4.73	4.35	5.11
RMSEA	0	61.06	60.19	61.93
RMSEA	50	32.69	31.85	33.53
RMSEA	70	3.69	3.35	4.03
SRMR	0	50.73	49.83	51.62
SRMR	50	28.29	27.48	29.09
SRMR	70	3.00	2.69	3.31

Table 2.: Mean percentage correct and its 95 confidence interval for each level of factor loadings

Method	Loadings	Mean	95% C.I. - Lower	95% C.I. - Upper
TEFI.vn	40	83.87	83.21	84.52
TEFI.vn	55	98.41	98.18	98.63
TEFI.vn	70	99.96	99.92	99.99
EFI.vn	40	82.56	81.88	83.24
EFI.vn	55	98.24	98.01	98.48
EFI.vn	70	99.92	99.88	99.97
SRMR.rel	40	74.85	74.07	75.62
SRMR.rel	55	96.39	96.05	96.72
SRMR.rel	70	95.17	94.79	95.56
EFI	40	74.73	73.96	75.51
EFI	55	89.92	89.39	90.46
EFI	70	93.40	92.96	93.84
RMSEA.rel	40	62.78	61.92	63.65
RMSEA.rel	55	92.38	91.91	92.86
RMSEA.rel	70	96.46	96.13	96.79
TLI.rel	40	68.13	67.30	68.97

TLI.rel	55	93.44	93.00	93.88
TLI.rel	70	93.32	92.87	93.76
CFI.rel	40	59.86	58.98	60.74
CFI.rel	55	90.79	90.27	91.31
CFI.rel	70	95.86	95.50	96.21
TLI	40	48.84	47.95	49.74
TLI	55	69.57	68.75	70.39
TLI	70	78.12	77.38	78.86
CFI	40	42.86	41.97	43.74
CFI	55	62.12	61.25	62.99
CFI	70	67.38	66.54	68.22
RMSEA	40	0.00	0.00	0.00
RMSEA	55	30.03	29.21	30.85
RMSEA	70	67.32	66.48	68.16
SRMR	40	0.00	0.00	0.00
SRMR	55	17.94	17.25	18.62
SRMR	70	64.06	63.20	64.92

The skew of the items also plays a role in the accuracy. When the skew is -2 (skew 1 in Figure 6), the overall accuracy is 72.98%, when the skew is -2 and 2 (skew 2 in Figure 6), the overall accuracy is 66.61%. For a skew ranging from -2 to 2, by increments of .5 (skew 3), the general accuracy is 74.63%, and when the skew is 2 (skew 4 in Figure 6) the overall accuracy is 75.30%.

TEFI.vn and EFI.vn stands out once again in the worst condition in terms of skew, presenting an accuracy of 92.64% and 91.80% for the skew -2 and 2, respectively, followed by SRMR.rel (80.56%), RMSEA.rel (77.78%), CFI.rel (75.66%), TLI.rel (75.44%) and EFI (74.26%).

Overall, TLI (47.53%), CFI (49.44%), RMSEA (46.81%), and SRMR (44.56%) presented low accuracies, and will not be presented in detail. Their low accuracy can be explained by the values of these fit indexes, that has a tendency to be in a range considered adequate by the cut-off points used ($\geq .95$ for CFI and TLI; $\leq .05$ for RMSEA and SRMR). Figure 7 shows the distribution of the fit indices values across all combination of four items into two factor for factor loadings of .40 and correlation between factors of .70. The values of the traditional indices indicates that all the incorrect structures could be considered as presenting a good fit, if the cut-off values are used. Note that all tested structures present CFI and TLI values higher than .95 and RMSEA and SRMR values lower than .05. Notably, the CFI and TLI values are maximized when the correct structure is specified, while RMSEA and SRMR values are minimized. This pattern helps to understand why using these indices as relative fit measures works much better than using their cut-off values.

Simulation 2: Shuffled, Underfactoring and Overfactoring Conditions

In terms of the overall performance, the fit indices can be separated into four groups (see Figure 8): (1) high accuracy (PC) in the four conditions tested (TEFI.vn), (2)

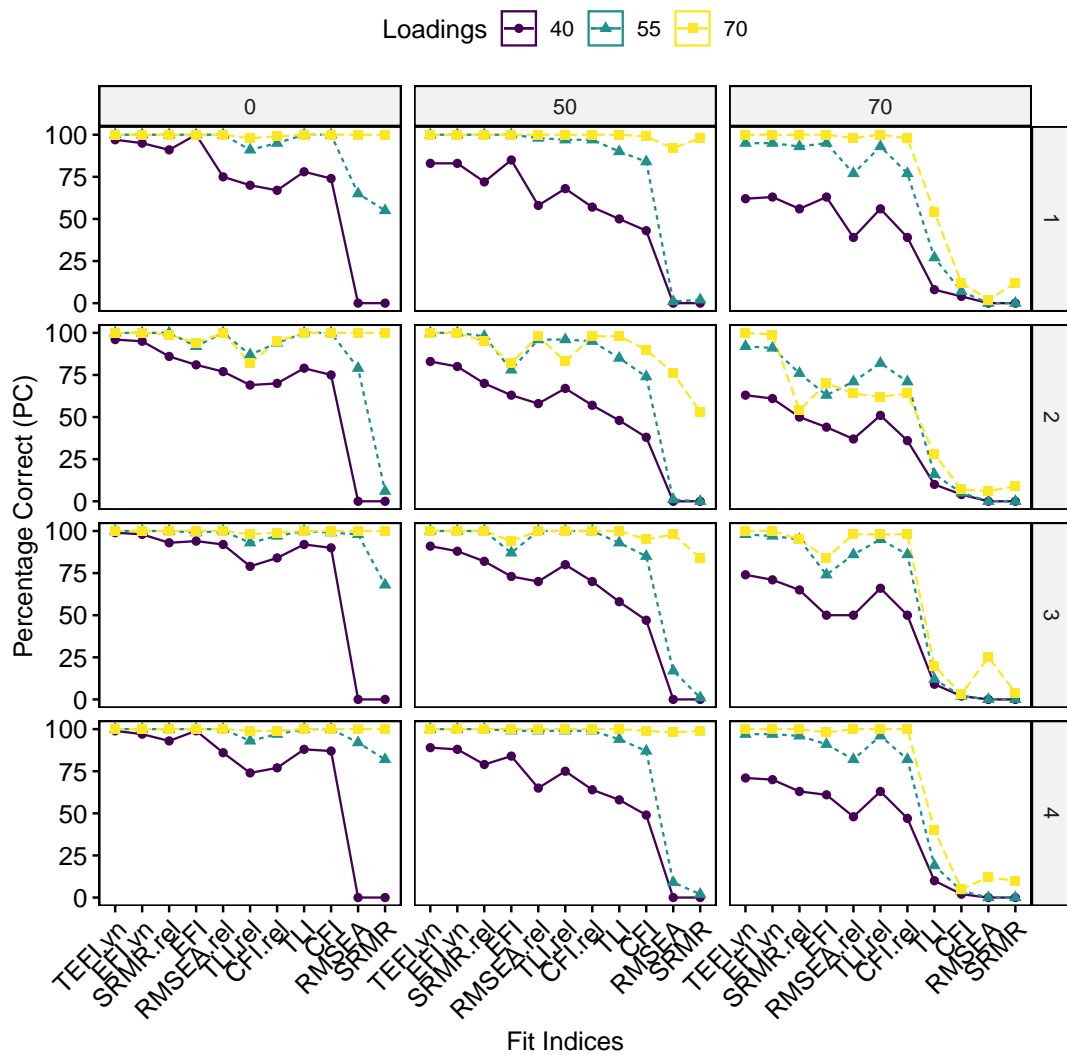


Figure 6. Percentage correct per levels of correlation (columns), skew (rows), loadings (color), and fit index

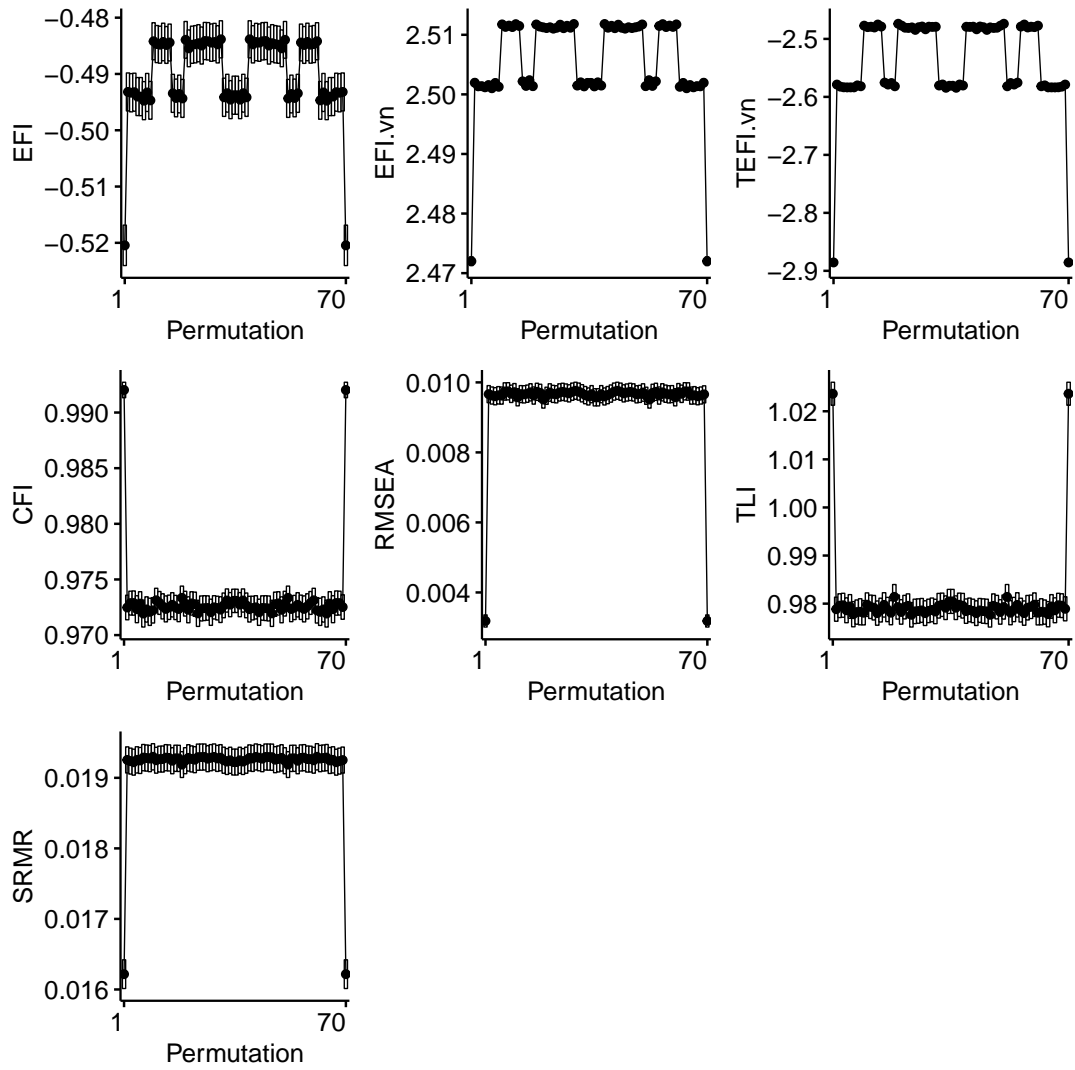


Figure 7. Distribution of the fit indices values across all combination of four items into two factors

high or moderately high PC in three of the four conditions (EFI, RMSEA.rel, CFI.rel, SRMR.rel, TLI.rel), (3) high PC in two conditions (EFI.vn), and (4) moderate or low PC across conditions (TLI, CFI, SRMR and RMSEA). TEFI.vn presented an overall accuracy of 92% (see Figure 8), followed by EFI (PC = 84%) and TLI.rel (PC = 82%). On the one hand, TEFI.vn presented very high PCs even in the most extreme overfactoring conditions (see Figure 9), especially with correlated factors (PC = 96.57%). On the other hand, TEFI.vn did not perform as well in the underfactoring condition with high factor loading and high interfactor correlation (PC = 58.95%). This problem does seem to vanish if the number of items is high (PC = 100%).

EFI presented a very high PC in the shuffled (99.68%) and underfactoring condition (99.98%), but only a moderately high PC in the overfactoring condition (85.72%) and a moderate PC in the more extreme overfactoring condition (52.45%). EFI.vn, by its turn, presented a very high PC in the shuffled (100%) and underfactoring condition (100%), but basically zero accuracy in both overfactoring conditions. The CFA indices used as relative fit measures (RMSEA.rel, CFI.rel, TLI.rel, and SRMR.rel) were stable across the shuffled, underfactoring, and overfactoring conditions, presenting PCs ranging from 98% and 100%. However, in the more extreme overfactoring conditions their PCs dropped to 36.94% for TLI.rel, 26.23% for RMSEA.rel, 15.84% for CFI.rel, and 0.50% for SRMR.rel. When applied using cut-off values only, the CFA fit measures presented low to moderate PCs in the shuffled, underfactoring, and overfactoring conditions as well as very low PC in the more extreme overfactoring condition (see Figure 8).

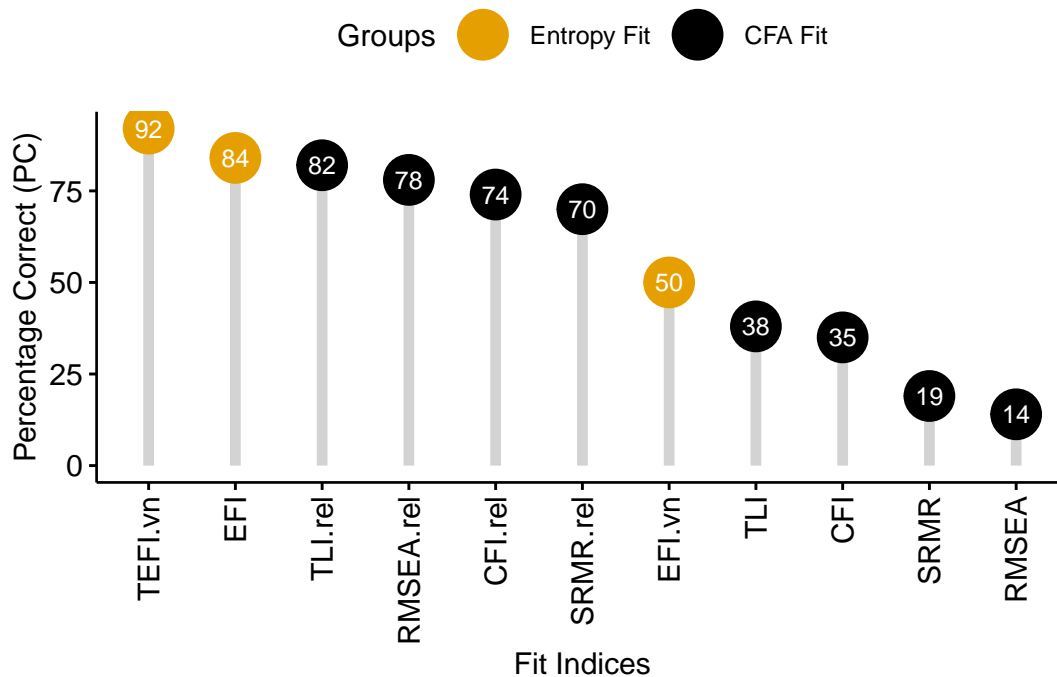


Figure 8. Mean percentage correct (PC) per fit index (simulation two)

Figure 9 shows a more detailed description of the performance of the fit indices across conditions. TEFI.vn presents a very high accuracy to detect the simulated number of dimensions with PC equal to (or very close to) 100%. Even in the most extreme

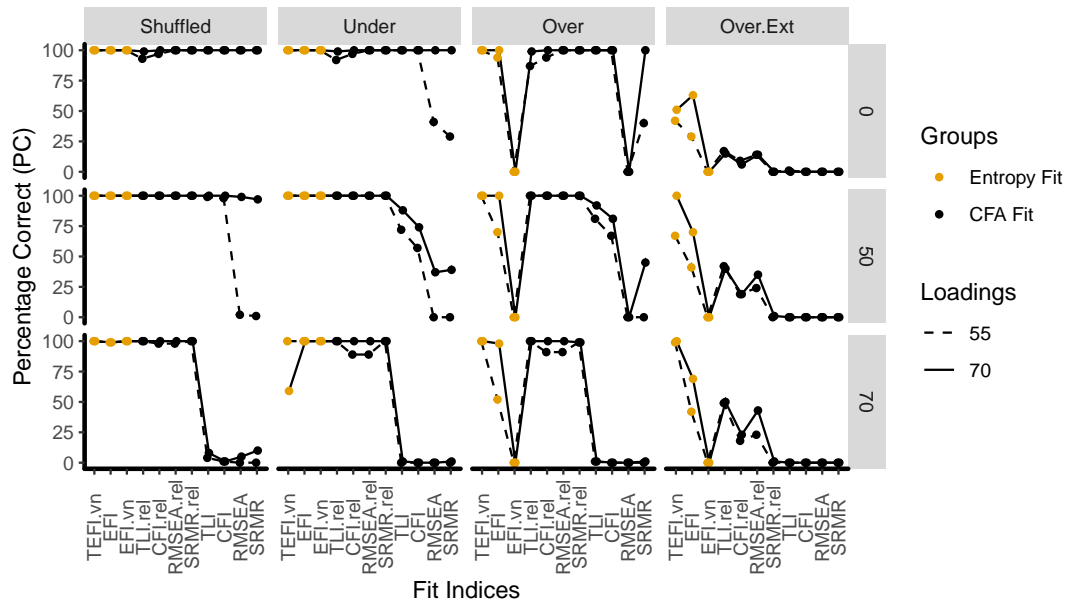


Figure 9. Mean percentage correct (PC) per fit index, in the shuffling, underfactoring and overfactoring conditions, for different levels of factor loadings and correlation between factors

overfactoring condition, TEFI.vn presents an impressive accuracy in which most of the other fit indices (except EFI) present moderately low to low accuracy (see Table 3). In this condition, and with factor loadings of .70, TEFI.vn presents a PC of 50.85% for inter-factor correlations of 0, 100% for inter-factor correlations of .50, and 100% for inter-factor correlations of .70 (see 9). Considering only the shuffled, underfactoring, and overfactoring conditions, the only situation where the accuracy of TEFI.vn is not very high is in the underfactoring condition with high interfactor correlation and high factor loading.

EFI, RMSEA.rel, CFI.rel, SRMR.rel, and TLI.rel presented very high PCs across conditions (shuffled, underfactoring or overfactoring) when the loadings were .70, irrespective of the correlation between factors, sample size, and number of items. In the more extreme overfactoring condition, however, RMSEA.rel, CFI.rel, SRMR.rel, and TLI.rel presented low accuracies while EFI presented a moderate accuracy. For factor loadings of .70 in the extreme overfactoring condition, EFI presented a PC of 62.95% for inter-factor correlations of 0, 69.75% for inter-factor correlations of 50, and 69.40% for inter-factor correlations of 70 (see Figure 9). Conversely, EFI presented PCs similar to the other CFA fit indices for factor loadings of .50, while EFI.vn failed almost in all overfactoring conditions.

In terms of the CFA fit indices, TLI.rel was the third most accurate fit index in the more extreme overfactoring condition. For factor loadings of .70 it presented a PC of 15.15% for inter-factor correlations of 0, 40.12% for inter-factor correlations of 50, and 50.31% for inter-factor correlations of 70.

Table 3 shows the percentage correct and its 95% confidence interval for each condition tested (shuffled, underfactoring, overfactoring and extreme overfactoring).

Table 3.: Mean percentage correct and its 95 confidence interval for each condition tested (shuffled, underfactoring, overfactoring and extreme overfactoring)

Method	Condition	Mean	95% C.I. - Lower	95% C.I. - Upper
TEFI.vn	Shuffled	100.00	100.00	100.00
TEFI.vn	Under	93.05	92.60	93.51
TEFI.vn	Over	100.00	100.00	100.00
TEFI.vn	Over.Ext	76.74	75.98	77.50
EFI	Shuffled	99.68	99.57	99.78
EFI	Under	99.98	99.95	100.00
EFI	Over	85.72	85.10	86.35
EFI	Over.Ext	52.45	51.56	53.34
EFI.vn	Shuffled	100.00	100.00	100.00
EFI.vn	Under	100.00	100.00	100.00
EFI.vn	Over	0.00	0.00	0.00
EFI.vn	Over.Ext	0.00	0.00	0.00
TLL.rel	Shuffled	99.26	99.11	99.41
TLL.rel	Under	99.20	99.04	99.36
TLL.rel	Over	99.03	98.85	99.20
TLL.rel	Over.Ext	36.94	36.08	37.81
CFI.rel	Shuffled	99.31	99.16	99.46
CFI.rel	Under	97.30	97.01	97.59
CFI.rel	Over	97.65	97.38	97.92
CFI.rel	Over.Ext	15.84	15.18	16.49
RMSEA.rel	Shuffled	99.60	99.48	99.71
RMSEA.rel	Under	97.60	97.32	97.87
RMSEA.rel	Over	97.97	97.72	98.22
RMSEA.rel	Over.Ext	26.23	25.44	27.01
SRMR.rel	Shuffled	100.00	100.00	100.00
SRMR.rel	Under	100.00	100.00	100.00
SRMR.rel	Over	99.66	99.56	99.77
SRMR.rel	Over.Ext	0.50	0.38	0.63
TLI	Shuffled	62.73	61.86	63.59
TLI	Under	49.98	49.08	50.87
TLI	Over	49.53	48.63	50.42
TLI	Over.Ext	0.18	0.11	0.26
CFI	Shuffled	60.38	59.50	61.25
CFI	Under	43.76	42.87	44.65
CFI	Over	43.84	42.95	44.73
CFI	Over.Ext	0.06	0.02	0.10
RMSEA	Shuffled	41.90	41.02	42.79
RMSEA	Under	19.85	19.13	20.56
RMSEA	Over	0.00	0.00	0.00

RMSEA	Over.Ext	0.00	0.00	0.00
SRMR	Shuffled	42.37	41.48	43.25
SRMR	Under	19.24	18.54	19.95
SRMR	Over	18.68	17.99	19.38
SRMR	Over.Ext	0.00	0.00	0.00

Applied Example:

The world health organization study on global ageing and adult health (SAGE)

This empirical example investigates the dimensionality structure of a large dataset from the World Health Organization (WHO) Study on global ageing and adult health (SAGE). Although being a longitudinal study focusing on adults aged 50 years and older, WHO’s SAGE also collects data from nationally representative adult samples (aged between 18 and 49 years) in China. In the current paper, only data from the first wave of data collection will be used.

The original sample is composed of 9,283 adults aged between 18 and 99 years (Mean = 60.53, SD = 11.91) with 53.40% being females. In terms of the highest level of education completed, 225.09% finished primary school, 18.22% high school (or equivalent), 27.96% secondary school, 21.82% less than primary school, and 6.7% have college/pre-university/university or post-graduate degree. The 9,283 participants answered all 52 questions related to broader domain labeled *activities and participation*, containing items related to six theoretical domains: (1) social, community and civic activities/involvement, (2) interpersonal relations, (3) mobility, (4) self-care, (5) domestic life and (6) personal history of work and employment.

The dataset was split into two random subsamples (training and testing set), each containing 50% of the data, for cross-validation. The results of both EGA estimations were compared to the theoretical dimensionality structure (six domains) briefly described above. The EFI, TEFI.vn, and traditional factor-analytic fit measures were computed for each structure (i.e. the structures estimated using EGA with GGM, EGAtmfg, plus the theoretical model) in the training and testing sample. The TEFI.vn index was used to select the optimal number of steps in the Walktrap algorithm (used in the EGA technique), comparing the dimensionality structure estimated via EGA using from three to 10 steps.

Table 4 shows the factor-analytic fit values of the theoretical model in the training (CFI = 0.90, TLI = 0.89, RMSEA = 0.04, SRMR = 0.06) and test set (CFI = 0.89, TLI = 0.89, RMSEA = 0.04, SRMR = 0.06), suggesting a poorer fit of this model compared with the other two empirical models.

The structure estimated via EGA presented the lowest TEFI.vn in both samples, compared with the theoretical structure and the structure estimated via EGAtmfg. Both EGA and EGAtmfg estimated five factors, but with a different composition of items per factor. The factor-analytic fit measures were basically identical for the factors estimated via EGA and EGAtmfg, while TEFI.vn was lower for the structure estimated using EGA in both samples (in the training and testing/validation sample).

Table 4.: Fit Measures for the SAGE dataset

Data	Model	Walktrap Steps	EFI	TEFI.vn	CFI	TLI	RMSEA	SRMR
Train	EGA	3	-5.03	-69.51	0.92	0.92	0.03	0.06
	EGAtmfg	3	-5.89	-55.95	0.93	0.93	0.03	0.06
	Theoretical Model		-4.89	-46.53	0.90	0.89	0.04	0.06
Test	EGA		-5.52	-68.61	0.92	0.92	0.03	0.06
	EGAtmfg		-5.99	-54.64	0.93	0.93	0.03	0.06
	Theoretical Model		-4.89	-46.01	0.90	0.89	0.04	0.06

Figure 10 shows the structure estimated via EGA and EGAtmfg, while table 4 presents the variables, the EGA and EGAtmfg factor numbers per variable, the description of the variables and the theoretical factor per variable. Factor one (red nodes in Figure 10) represents a *socialization* factor, with items such as *Q6001* (attending public meeting), *Q6005* (had friends over to your home) and *Q6009* (went out of house to see friends or relatives), among others. Factor two (yellow/green-ish nodes) contains items related to work and employment (e.g. *Q6020*: Do/did you usually work throughout the year, or do/did you work seasonally, or only once in a while for your main job?) and to the participation in the political life (e.g. *Q6021*: Lots of people find it difficult to get out and vote. Did you vote in the last state/national/presidential election?). Factor three (light green nodes) contains items related to overall difficulty in the past 30 days with self-care (e.g., *Q2036*: In the last 30 days, how much difficulty did you have bathing/washing your whole body?), domestic work (e.g., *Q2039*: In the last 30 days, how much difficulty did you have day to day work?), mobility (e.g., *Q2026*: In the last 30 days, how much difficulty did you have walking 100 meters), and with engaging in social events (*Q2033*: In the last 30 days, how much difficulty did you have joining community activities?). Factor four (blue nodes) presents items related to overall difficulty in daily activities (work or household activities, moving around, doing vigorous activities, bathing, taking care of appearance and staying by yourself). Factor five, by its turn, contains items related to overall difficulty with interpersonal relations and activities.

Discussion

Reducing the dimensions of multivariate data has proven to be an extremely worthwhile task in the study of behavior. This has generally taken the form of procedures inferring the number and nature of unmeasured (latent) variables (e.g., factors) that can account for the calculated relations among the observed variables. Willingness to ascribe causal properties to these latent variables varies between investigators and is not the primary concern of this article; however, accurately estimating the number of dimensions (e.g., latent variables) is our primary concern and often the main concern of investigators.

For many researchers with a causal latent variable perspective, the relations between latent variables can provide the generality needed to articulate useful lawfulness. An example used by Nesselroade and Molenaar (2016) concerned the *frustration produces*

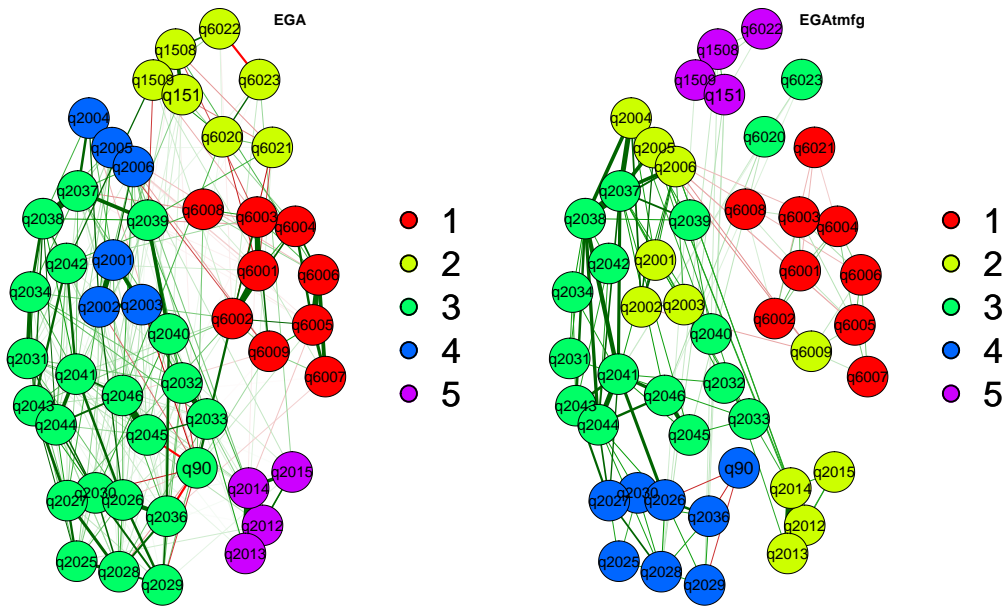


Figure 10. Structure of the SAGE data using EGA

aggression hypothesis. Some students, frustrated because they cannot solve a difficult mathematics problem, might curse the teacher whereas other students might slam down their books and walk out of class and still others might slap at a fellow student nearby. To say that “*inability to solve a difficult math problem leads to slapping a fellow student*” might hold for some students but it does not have generality whereas at the latent variable level “*frustration produces aggression*” covers all three situations above. Obviously, the properties of frustration and aggression must be inferred from the observable variables such as cursing, walking out, and slapping, but to build lawfulness that has generality, latent variables are key. The relations among latent variables are central to constructing or inferring mechanisms that help build common explanations of behavior. Extracting the most out of one’s multivariate data relies heavily on model fitting that best estimates the dimensionality and nature of the latent variables. Therefore, powerful methods for fitting models to data and judging the goodness of that fit are indispensable.

It is common practice among applied researchers and methodologists to use traditional fit indices from factor analysis and structural equation modeling to assess the factor solutions of their data because it provides useful diagnostic information, but also enables the comparison of different dimensionality structures. As pointed out earlier, there are only a few simulation studies investigating the accuracy of traditional fit indices (such as CFI and RMSEA) to estimate the correct number of factors (Clark & Bowles, 2018; Frazier & Youngstrom, 2007; Garrido et al., 2016; Yang & Xia, 2015) and, in general, they suggest that the traditional fit indices should be used with caution, especially when analyzing dichotomous data. Clark and Bowles (2018), for example, concluded that the common fit statistics cut-offs are, at best, imperfect tools for guiding decisions regarding dimensionality. The final suggestion is to use fit

indices in conjunction with other methods for assessing dimensionality. This suggestion is perhaps best exemplified by combining EGA with the new fit indices presented in the current paper (especially TEFI.vn and EFI) to obtain optimal dimension identification.

The accuracy of the entropy fit measures in correctly identifying the number of simulated factors was compared to RMSEA, CFI, TLI, and SRMR in two Monte Carlo simulations. In the first simulation, two latent factors with four variables each were simulated and the fit indices were used to check the fit of 70 different combinations of variables within the two factors. The goal of the first simulation was to identify the accuracy of the fit indices in detecting the correct simulated structure, given that the number of factors is correct but the placement of items per factor isn't. The results show that TEFI.vn and EFI.vn are the most accurate fit measures, followed by SRMR (when used as a relative measure of fit) and EFI. In general, the RMSEA/SRMR cut-offs presented very low accuracies in almost all conditions tested, except when the factor loadings were high and the interfactor correlation was zero or .50. Even in this favorable condition, the accuracy dropped for highly skewed items (skew 2; skew -2 and 2). For conditions with high interfactor correlation (.70), the accuracy of the traditional fit indices, using the cut-off values only, was very low. When used as relative measures of fit, the traditional fit indices (CFI, TLI, RMSEA and SRMR) presented high accuracies, comparable to TEFI.vn, EFI.vn and EFI, in almost all conditions.

In the second simulation, we investigated the accuracy of the fit indices to identify the correct structure compared to shuffled, underfactored, and overfactored conditions. TEFI.vn was the most accurate fit index to correctly identify the correct dimensionality structure. TEFI.vn presented a very high accuracy even in the most extreme overfactoring condition, a condition in which all the other fit indices (except EFI) presented moderately low or low accuracies. EFI presented a very high accuracy in the shuffled and underfactor conditions, a moderately high accuracy in the overfactoring condition, and was the second most accurate fit measure in the most extreme overfactoring condition (i.e. when one factor is split into two factors, without any items from other factors).

EFI.vn was also very accurate in the shuffled and underfactoring conditions, but completely failed in the overfactoring conditions. This result makes sense, since Von Neumann's entropy was developed to understand entangled quantum states. In the current simulation, the underfactor condition can be thought as reflecting mixed "states", since two simulated factors are merged as one single factor. Therefore, the EFI.vn value increases when two factors are merged into one relative to the EFI.vn value of the original simulated structure. However, the EFI.vn measure could not detect the correct simulated structure when one factor is split into two factors. When a state is not mixed but instead separated, then EFI.vn value remains low because the states are still reflecting homogeneous states. This may also be a consequence of the strategy used to penalize the number of factors. Future studies should investigate different penalization methods to verify if it is possible to increase the accuracy of EFI.vn in the overfactoring conditions. A possible explanation to the difference in accuracy between EFI.vn and the TEFI.vn is that the former replaced the joint entropy of the factors by the total entropy of the correlation matrix, which can be more robust in conditions where the partition of the multidimensional space is plausible but incorrect (i.e., extreme overfactoring condition).

The traditional fit indices, when used as absolute measures of fit, presented moderate to high accuracies in the shuffled and underfactoring conditions with interfactor

correlations of zero (orthogonal structures). Notably, the accuracies dropped with the increase of interfactor correlations becoming very low for interfactor correlations of .70. A very similar pattern occurred for the overfactoring condition. However, when the overfactoring condition was more extreme (one factor split into two pure factors), the cut-off values presented low accuracies irrespective of the interfactor correlation and factor loadings. Finally, when used as relative measures, the traditional fit indices presented very high accuracies for all conditions, except for the more extreme overfactoring condition. These measures, which are usually thought of as absolute are actually not. Instead, our results show that they are better thought of as relative measures. Therefore, our results suggest that these measures should not be used as absolute but rather relative fit measures.

In sum, the current paper corroborates previous simulation studies, showing that the traditional factor-analytic fit indices should be avoided for dimensionality assessment purposes and not used as absolute measures of fit (based on cut-off values). They are, however, very accurate when used as relative measures of fit specifically for the shuffled, underfactoring, and non-extreme overfactoring conditions. Also, the three new fit indices (EFI, EFI.vn, and TEFI.vn) were accurate in identifying the simulated number of factors and the respective placement of items per factor in the shuffled and underfactoring conditions. Notably, only TEFI.vn presented a high accuracy in both overfactoring conditions. These results represent an important advance in the area of measures of fit for structural and dimensionality analysis.

Finally, an empirical example was presented, using data from the WHO's Study on Global Ageing and Adult Health (SAGE) from China. Exploratory graph analysis was used to estimate the dimensionality structure using the GGM and the TMFG approaches, and their fit were compared to the fit of a theoretical model, in two random subsamples. The *TEFI.vn* index pointed to the EGA (GGM) structure (five factors) as the structure with the best fit to the data (compared to the theoretical model and to the structure estimated via EGAtmfg) in both samples (training and testing/validation). Basically, CFI, TLI, RMSEA and SRMR could not differentiate between the EGA and the EGAtmfg structures, both presenting five factors but with a different composition of items per factor.

Limitations

The limitations of the current paper can be summarized in two main aspects. The first one related to the data simulation approach used, which did not include correlated residuals (minor factors) at the population level. Moreover, only dichotomous data was used in the simulation. Future studies would bring new light into this area by incorporating other types of data (e.g., polytomous and mixed type: continuous, dichotomous, and polytomous). Second, the entropy fit indices only work for multidimensional factors (i.e., structures with at least two factors) presenting a simple structure (one item linked only to one factor). EGA uses the Walktrap algorithm, which is a technique that identifies communities in weighted networks where one node (or variable/item) is part of only one community (or factor). The entropy fit indices were developed as a way to check the fit of different dimensionality structures estimated via EGA (and to compare them to theoretical structures). Future studies should extend the entropy fit measures to accommodate unidimensional structures and to allow cross-loadings. Moreover, future research should examine the extension of the EFIs to other

structural equation modeling structures, which would greatly expand their utility to a larger number of researchers.

Conclusion

Our main conclusion is that EFI and TEFI.vn are accurate fit measures that can be used to check the fit of different dimensionality structures. They are relative measures of fit that present higher accuracies in detecting the correct dimensionality structure than traditional fit measures used in exploratory factor analysis and SEM. However, contrary to CFI, RMSEA, TLI and SRMR, the entropy fit measures cannot be used to evaluate the absolute fit of a model. EFI and TEFI.vn were developed to compare alternative dimensionality solutions, evaluating the partitioning of a multidimensional space in terms of its uncertainty or unstableness. As with the K-function, EFI and TEFI.vn are measures of entropy reduction by partitioning, but unlike the latter, the new fit measures do not decrease with the increase in the number of factors.

References

- Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, *34*(1), 122–148. Retrieved from https://www.jstor.org/stable/2991288?seq=1#metadata_info_tab_contents
- Antos, A., & Kontoyiannis, I. (2001). Convergence properties of functional estimates for discrete distributions. *Random Structures & Algorithms*, *19*(3-4), 163–193. <https://doi.org/https://doi.org/10.1002/rsa.10019>
- Asparouhov, T., & Muthén, B. (2018). *SRMR in mplus*. MPlus. Retrieved from <http://www.statmodel.com/download/SRMR2.pdf>
- Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 87–101. <https://doi.org/10.1080/10705511.2014.934850>
- Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in cfa. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302/_2
- Beierl, E. T., Bühner, M., & Heene, M. (2018). Is that measure really one-dimensional? Nuisance parameters can mask severe model misspecification when assessing factorial validity. *Methodology*, *14*(4), 188–196. <https://doi.org/10.1027/1614-2241/a000158>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. <https://doi.org/doi:10.1037/0033-2909.107.2.238>
- Bollmann, S., Heene, M., Küchenhoff, H., & Bühner, M. (2015). *What can the real world do for simulation studies? A comparison of exploratory methods*. LMU. Retrieved from Department%20of%20Statistics,%20University%20of%20Munich:%20https://epub.ub.uni-muenchen.de/24518/
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. <https://doi.org/https://doi.org/10.1177/0049124192021002005>
- Campbell-Sills, L., Liverant, G. I., & Brown, T. A. (2004). Psychometric evaluation of the behavioral inhibition/behavioral activation scales in a large sample of outpatients with anxiety and mood disorders. *Psychological Assessment*, *16*(3), 244. <https://doi.org/10.1037/1040-3590.16.3.244>
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10
- Cellucci, C. J., Albano, A. M., & Rapp, P. E. (2005). Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms. *Physical Review E*, *71*(6), 066208. <https://doi.org/10.1103/PhysRevE.71.066208>
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical

- evaluation of the use of fixed cutoff points in rmsea test statistic in structural equation models. *Sociological Methods & Research*, 36(4), 462–494. <https://doi.org/10.1177/0049124108314720>
- Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3), 759–771. Retrieved from <https://www.jstor.org/stable/20441500>
- Christensen, A. P., Cotter, K. N., & Silvia, P. J. (2019). Reopening openness to experience: A network analysis of four openness to experience inventories. *Journal of Personality Assessment*, 101(6), 574–588. <https://doi.org/10.1080/00223891.2018.1467428>
- Clark, D. A., & Bowles, R. P. (2018). Model fit and item factor analysis: Overfactoring, underfactoring, and a program to guide interpretation. *Multivariate Behavioral Research*, 53(4), 544–558. <https://doi.org/10.1080/00273171.2018.1461058>
- Comrey, A. L., & Lee, H. B. (2016). *A first course in factor analysis*. New York: Routledge.
- 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>
- DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness? *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 453–466. <https://doi.org/10.1080/10705511.2017.1390394>
- Epskamp, M., S. (2018). Network psychometrics. In B. Irwing Paul (Ed.), *The wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 953–986). John Wiley & Sons Ltd. <https://doi.org/10.1002/9781118489772.ch30>
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48(4), 1–18. <https://doi.org/10.18637/jss.v048.i04>
- Epskamp, S., & Fried, E. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617–634. <https://doi.org/10.1037/met0000167>
- Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network pschometrics: Combining network and latent variable models. *Psychometrika*, 82(4), 904–927. <https://doi.org/10.1007/s11336-017-9557-x>
- Ferenci, T., & Kovács, L. (2014). Using total correlation to discover related clusters of clinical chemistry parameters. In *12th international symposium on intelligent systems and informatics* (pp. 49–54). IEEE. Retrieved from <https://core.ac.uk/download/pdf/42930601.pdf>
- Foygel, R., & Drton, M. (2010). Extended bayesian information criteria for gaussian

- graphical models. In *Proceedings of the 23rd international conference on neural information processing systems - volume 1* (Vol. 1, pp. 604–612). Vancouver, Canada.
- Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence*, *35*(2), 169–182. <https://doi.org/10.1016/j.intell.2006.07.002>
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. <https://doi.org/10.1093/biostatistics/kxm045>
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of velicer’s minimum average partial factor retention method with categorical variables. *Educational and Psychological Measurement*, *71*(3), 551–570. <https://doi.org/10.1177/0013164410389489>
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at horn’s parallel analysis with ordinal variables. *Psychological Methods*, *18*(4), 454–74. <https://doi.org/10.1037/a0030005>
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via monte carlo simulation. *Psychological Methods*, *21*(1), 93–111. <https://doi.org/10.1037/met0000064>
- Golino, H., & Christensen, A. P. (2019). *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), e0174035. <https://doi.org/10.1371/journal.pone.0174035>
- Golino, H., Shi, D., Garrido, L. E., Christensen, A. P., Nieto, M. D., Sadana, R., ... 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), 292–230. <https://doi.org/10.1037/met0000255>
- Hall, B. C. (2013). *Quantum theory for mathematicians* (Vol. 267). Springer.
- Harte, J., & Newman, E. A. (2014). Maximum information entropy: A foundation for ecological theory. *Trends in Ecology & Evolution*, *29*(7), 384–389. <https://doi.org/10.1016/j.tree.2014.04.009>
- 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(2), 191–205. <https://doi.org/10.1177/1094428104263675>
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16(3), 319–336. <https://doi.org/10.1037/a0024917>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185.
- Hu, L.-T., & Bentler, P. M. (1999a). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hu, L.-T., & Bentler, P. M. (1999b). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Joreskog, K. G., & Sorbom, D. (1981). *LISREL 5: Analysis of linear structural relationships by maximum likelihood and least squares methods;[user's guide]*. University of Uppsala.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on ackerman, beier, and boyle (2005). *Psychological Bulletin*, 131, 66–77. <https://doi.org/10.1037/0033-2909.131.1.66>
- Kassambara, A. (2018). *Ggpubr: 'Ggplot2' based publication ready plots*. Retrieved from <https://CRAN.R-project.org/package=ggpubr>
- 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>
- Laub, A. J. (2005). *Kronecker products* (Vol. 91). SIAM: Society for Industrial; Applied Mathematics.
- Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Oxford: Clarendon Press.
- Li, C. (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>
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing hu and bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
- Massara, G. P., Di Matteo, T., & Aste, T. (2016). Network filtering for big data: Triangulated maximally filtered graph. *Journal of Complex Networks*, 5(2), 161–178. <https://doi.org/10.1093/comnet/cnw015>

- McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment*, *100*(1), 43–52. <https://doi.org/10.1080/00223891.2017.1281286>
- Meade, A. W. (2008). Power of afi's to detect cfa model misfit. In *Paper presented at the 23th annual conference of the society for industrial and organizational psychology*. San Francisco, CA. Retrieved from pdfs.semanticscholar.org/a23c/45ca18db70125a9a0ad983926513d40fa32b.pdf
- Mezard, M., & Montanari, A. (2009). *Information, physics, and computation*. Oxford University Press.
- Moreau, P.-A., Toninelli, E., Gregory, T., Aspden, R. S., Morris, P. A., & Padgett, M. J. (2019). Imaging bell-type nonlocal behavior. *Science Advances*, *5*(7), eaaw2563. <https://doi.org/10.1126/sciadv.aaw2563>
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide (version 7)*. Los Angeles, CA: Author. MPlus.
- Nesselroade, J. R., & Molenaar, P. C. (2016). Some behavioral science measurement concerns and proposals. *Multivariate Behavioral Research*, *51*(2-3), 396–412. <https://doi.org/10.1080/00273171.2015.1050481>
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*. <https://doi.org/10.1103/PhysRevE.69.066133>
- Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Computation*, *15*(6), 1191–1253. <https://doi.org/10.1162/089976603321780272>
- Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. In P. Yolum, T. Güngör, F. Gürgen, & C. Özturan (Eds.), *Computer and information sciences - iscis 2005* (pp. 284–293). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11569596_31
- Preskill, J. (2018). Quantum shannon entropy. In J. Preskill (Ed.), *Quantum information* (p. 94). Cambridge University Press. Retrieved from <https://arxiv.org/pdf/1604.07450.pdf>
- Raiche, G., Walls, T. A., Magis, D., Riopel, M., & Blais, J.-G. (2013). Non-graphical solutions for cattell's scree test. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *9*(1), 23–29. <https://doi.org/10.1027/1614-2241/a000051>
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- 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>
- Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial

- structure. *Psychological Assessment*, 24(2), 282–292. <https://doi.org/10.1037/a0025697>
- Sanne, B., Torsheim, T., Heiervang, E., & Stormark, K. M. (2009). The strengths and difficulties questionnaire in the bergen child study: A conceptually and methodically motivated structural analysis. *Psychological Assessment*, 21(3), 352. <https://doi.org/10.1037/a0016317>
- Santos, L., Vagos, P., & Rijo, D. (2018). Dimensionality and measurement invariance of a brief form of the young schema questionnaire for adolescents. *Journal of Child and Family Studies*, 27(7), 2100–2111. <https://doi.org/10.1007/s10826-018-1050-3>
- Savalei, V. (2012). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. *Educational and Psychological Measurement*, 72(6), 910–932. <https://doi.org/10.1177/0013164412452564>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.
- Steiger, J. H. (1980). Statistically based tests for the number of common factors. In *The annual meeting of the psychometric society*. Iowa city, ia.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- 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>
- Tsuruyama, T. (2018). Entropy in cell biology: Information thermodynamics of a binary code and szilard engine chain model of signal transduction. *Entropy*, 20(8), 617. <https://doi.org/10.3390/e20080617>
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10. <https://doi.org/10.1007/BF02291170>
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321–327. <https://doi.org/10.1007/BF02293557>
- Ventimiglia, M., & MacDonald, D. A. (2012). An examination of the factorial dimensionality of the marlowe crowne social desirability scale. *Personality and Individual Differences*, 52(4), 487–491. <https://doi.org/10.1016/j.paid.2011.11.016>
- Volkenstein, M. V. (2009). *Entropy and information* (1st ed.). Birkhauser.
- Von Neumann, J. (1927). Wahrscheinlichkeitstheoretischer aufbau der

- quantenmechanik. *Nachrichten von Der Gesellschaft Der Wissenschaften Zu Göttingen, Mathematisch-Physikalische Klasse*, 1927, 245–272.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236–244. <https://doi.org/10.2307/2282967>
- Watanabe, H. (2000). A new view of the formal entropy as a measure of interdependence and its application to pattern recognition. In *SMC 2000 conference proceedings. 2000 IEEE international conference on systems, man and cybernetics: "Cybernetics evolving to systems, humans, organizations, and their complex interactions"* (Vol. 4, pp. 2827–2832). <https://doi.org/10.1109/ICSMC.2000.884426>
- Watanabe, H. (2001). Clustering as average entropy minimization and its application to structure analysis of complex systems. In *"2001 IEEE international conference on systems, man and cybernetics": "E-systems and e-man for cybernetics in cyberspace"* (Vol. 4, pp. 2408–2414). <https://doi.org/10.1109/ICSMC.2001.972918>
- Watanabe, S. (1939). Über die anwendung thermodynamischer begriffe auf den normalzustand des atomkerns. *Zeitschrift Für Physik*, 113(7-8), 482–513.
- Watanabe, S. (1960). Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, 4(1), 66–82. <https://doi.org/10.1147/rd.41.0066>
- Watanabe, S. (1969). *Knowing and guessing: A formal and quantitative study*. New York: John Wiley & Sons.
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <http://ggplot2.org>
- Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28(3), 263–311. https://doi.org/10.1207/s15327906mbr2803_1
- Wiener, N. (1961). *Cybernetics or control and communication in the animal and the machine* (Vol. 25). MIT press.
- Wihler, T. P., Bessire, B., & Stefanov, A. (2014). Computing the entropy of a large matrix. *Journal of Physics A: Mathematical and Theoretical*, 47(24), 245201. <https://doi.org/10.1088/1751-8113/47/24/245201>
- Xia, Y., & Yang, Y. (2018). The influence of number of categories and threshold values on fit indices in structural equation modeling with ordered categorical data. *Multivariate Behavioral Research*, 53(5), 731–755. <https://doi.org/10.1080/00273171.2018.1480346>
- 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>
- Yeung, R. W. (2008). *Information theory and network coding*. Springer Science & Business Media.

- Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Retrieved from <https://pdfs.semanticscholar.org/7a22/ae22553f78582fc61c6cab4567d36998293b.pdf>
- Zhao, C., Yang, G.-w., Hung, W. N., & Li, X.-y. (2015). A multipartite entanglement measure based on coefficient matrices. *Quantum Information Processing*, 14(8), 2861–2881. <https://doi.org/https://doi.org/10.1007/s11128-015-1023-z>
- Zurek, W. H. (2018). *Complexity, entropy and the physics of information*. CRC Press.