Generalized Total Entropy Fit Index: A new fit index to compare bifactor and correlated factor structures in SEM and network psychometrics

Hudson Golino[1], Marcos Jimenez[2], Luis Eduardo Garrido[3], and & Alexander P. Christensen[4]

[1] University of Virginia

[2] Universidad Autónoma de Madrid

[3] Pontificia Universidad Católica Madre y Maestra

[4] Vanderbilt University

Author Note

The authors made the following contributions. Hudson Golino: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft Preparation, Writing - Review & Editing; Marcos Jimenez: Methodology, Writing - Original Draft Preparation, Writing - Review & Editing; Luis Eduardo Garrido: Methodology, Writing - Original Draft Preparation, Writing - Review & Editing; Alexander P. Christensen: Methodology, Writing - Original Draft Preparation, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Hudson Golino, Postal address. E-mail: hfg9s@virginia.edu

Abstract

The Generalized Total Entropy Fit Index (GenTEFI) is introduced as a new fit index for comparing the dimensionality of correlated traits and hierarchical/bifactor structures in Structural Equation Modeling (SEM) and Network Psychometrics. This study addresses limitations in recent simulation studies that focus solely on discrepancy due to estimation, by incorporating discrepancy due to approximation and population error. Three Monte Carlo simulations were conducted to assess the accuracy of GenTEFI and traditional fit indices (CFI, RMSEA, SRMR, BIC, AIC) in identifying correct model structures at both sample and population levels (correlated traits vs. bifactor, bifactor vs. correlated traits, bifactor vs. bifactor). Results show that GenTEFI is the only fit index that adequately differentiates between correlated traits and bifactor structures in both sample and population contexts, regardless of the true data generation mechanism. GenTEFI demonstrated high balanced accuracy (94-95%) across all conditions tested, outperforming traditional fit indices like BIC, AIC, RMSEA, and SRMR, when the data comes from a correlated traits model. When the data comes from a bifactor model, or when comparing different bifactor structures, GenTEFI performed comparably to traditional indices. Two empirical examples illustrate the practical application of GenTEFI in model selection for attachment styles and misinformation susceptibility. This study highlights the potential of GenTEFI to enhance the validity and reliability of measurement instruments in psychology, offering researchers a powerful new tool for understanding the structure of psychological constructs.

Keywords: keywords

Word count: X

Generalized Total Entropy Fit Index: A new fit index to compare bifactor and correlated factor structures in SEM and network psychometrics

Introduction

Psychometric research has increasingly focused on understanding the complex structure of psychological constructs using advanced modeling techniques. The bifactor model, which allows for the estimation of both general and group factors, has gained popularity among researchers due to its ability to provide a more nuanced representation of the underlying structure of the data. However, recent simulation studies have raised concerns about the potential bias of fit indices when comparing correlated traits and bifactor models, highlighting the need for caution in model selection (Greene et al., 2019; Morgan, Hodge, Wells, & Watkins, 2015; Murray & Johnson, 2013).

Fit indices, such as the comparative fit index (*CFI*: Bentler, 1990) and the root mean square error of approximation (*RMSEA*: Steiger, 1980), are commonly used to assess the adequacy of competing models and determine which model best fits the data. These indices are widely relied upon to justify the selection of one model over another, particularly in the context of comparing correlated traits and bifactor models (Morgan et al., 2015). However, simulation studies have revealed that fit indices may be biased in favor of bifactor models, even when the data generating mechanism is a correlated traits model (Greene et al., 2019; Morgan et al., 2015) or when the data follow random patterns (Bonifay & Cai, 2017). This bias is particularly concerning because researchers rely on model fit to determine the most appropriate model.

Cucina and Byle (2017a), for example, showed the bifactor model was the model presenting better fit across 166 comparisons (90% of the comparisons), leading to the recommendation for researchers to consider using bifactor models in intelligence batteries. Another representative example is the work of Snyder, Young, and Hankin (2017) that showed the bifactor structure presented a better fit than the unidimensional and the two correlated traits model, therefore recommending the former as the the best representation of psychopathology symptoms (the p-factor of psychopathology). The tendency of bifactor models to overfit the data, especially in the presence of model misspecifications or complexities in the true data-generating process, further complicates the interpretation and generalizability of the findings of research investigating the structural organization of variables in psychology (Bonifay, Lane, & Reise, 2017).

Figure 1 illustrates the issue by showing a correlated traits (left) and a bifactor model (right) fitted to data generated from a correlated traits model[1]. Thickness in Figure 1 represents loading size (items with higher loadings have thicker edges in the plot).

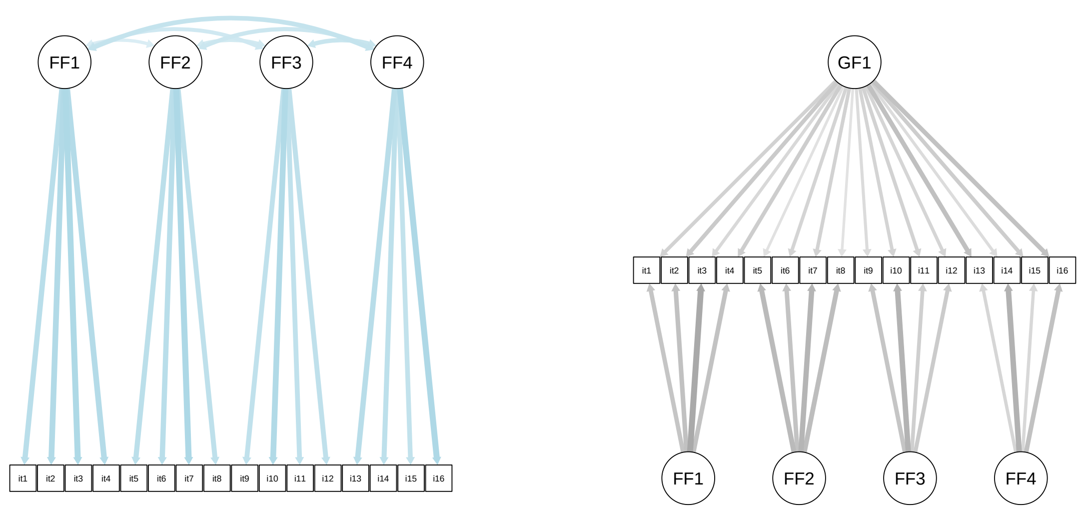


Figure 1
Fitting a Correlated Traits Model and a Bifactor Model to Data Generated from Correlated Traits.

The correlated traits model had a CFI of 0.92, a RMSEA of 0.03 and a SRMR of 0.03, while the bifactor model presented a CFI of 0.93, a RMSEA of 0.03 and a SRMR of 0.03. Despite being the correct model, the correlated traits model ($\chi^2 = 212.00$) presented a

[1] The data generation mechanism has four variables per factor, moderate loadings (ranging from 0.45 to 0.55), and interfactor correlation of 0.30. Sample size: 1,000 observations.

significantly worse fit to the data than the bifactor model ($\chi^2 = 192.46$, $\Delta\chi^2 = 19.54$, $\Delta df = 10$, $p = 0.03$).

In a recent study, Kan, Psychogyiopoulos, Groot, Jonge, and Ten Hove (2024) cautioned against using approximate fit indices (such as CFI, RMSEA) as relative measures when comparing models, as they may be biased towards favoring the more parameterized bifactor models. The authors recommend using relative fit indices like the Bayesian information criterion (BIC) and the Akaike information criterion (AIC) for model selection between bifactor other types of models (Kan et al., 2024).

To address the potential bias of bifactor models, studies have emphasized the need for a more comprehensive approach to model evaluation and selection. Rather than relying solely on fit indices, researchers have been urged to consider a range of factors, including substantive theory, factor strength indices, and the interpretability of the resulting factor scores (Reise, Scheines, Widaman, & Haviland, 2013; Rodriguez, Reise, & Haviland, 2016). The usual argument is that by adopting a more nuanced and critical approach to the application and interpretation of bifactor models, researchers can mitigate the potential bias and ensure that their conclusions are grounded in both statistical and substantive considerations.

Despite the significant practical and theoretical implications of overfitting bifactor models, especially in contrast to correlated traits structures, it is remarkable that the primary concern—the efficacy of fit indices—has often been overlooked. Many studies emphasize the need for parsimony in model comparison, cautioning against exclusive dependence on fit indices. This dependence raises a critical question: if fit indices are not reliable for assessing model-data congruence, what purpose do they serve? The prevalent reliance on these indices without addressing their limitations underlines a significant gap in the field.

Instead of tackling the problem of traditional fit indices being biased toward bifactor models, suggesting that *new fit indices are necessary*, the field has adopted either an "evasion" approach (e.g., "researchers shouldn't rely solely on fit index, but use theory or a parsimonious approach") or a "bifactor avoidance" position (e.g., "bifactor models should not be used because they overfit the data"). Neither positions addresses the problem, but both lead to a lack of engagement with critical or challenging aspects of the subject, potentially hindering scientific progress. The field has historically prioritized tradition, and may have overlooked or underestimated the need for new fit indices that are not based on the distance, difference, or equivalence between a model-implied covariance matrix and the covariance matrix of observed variables (like the traditional fit indices CFI, RMSEA, and SRMR).

Limitations of Recent Simulation Studies

Recent simulation studies investigating the performance of fit indices when comparing correlated traits and bifactor models have adopted strategies that limit their generalizability: they focus solely on discrepancy due to estimation, which quantifies sampling error or the discrepancy between population and sample estimates (Cudeck & Henly, 1991; Preacher, Zhang, Kim, & Mels, 2013). Cudeck and Henly (1991) defined four types of discrepancy, described below and represented in Figure 2:

1. Discrepancy due to Approximation (DA): This represents the difference between the population covariance matrix ($\boldsymbol{\Sigma}_0$) and the model's implied covariance matrix in the population ($\tilde{\boldsymbol{\Sigma}}_0$). DA is a measure of the model's lack of fit in the population, independent of sample size and sampling variability. It quantifies the degree of model misspecification or the extent to which the model deviates from the true data-generating process. Minimizing DA is equivalent to maximizing the model's verisimilitude or proximity to the truth. It quantifies population error or model misspecification, independent of sampling variability.

2. Discrepancy due to Estimation (DE): This represents the sampling variability or the difference between the model's implied covariance matrix in the population ($\tilde{\boldsymbol{\Sigma}}_0$) and

the model's implied covariance matrix in the sample (or model prediction: $\hat{\mathbf{\Sigma}}$). DE arises because model parameters are estimated from a finite sample rather than the population. As sample size increases, DE decreases, and $\hat{\mathbf{\Sigma}}$ approaches $\tilde{\mathbf{\Sigma}}_0$. It quantifies sampling error or the discrepancy between population and sample estimates.

3. Sample Discrepancy (SD): Refers to the discrepancy between observed data (S) and a model's predictions ($\hat{\mathbf{\Sigma}}$).

4. Overall Discrepancy (OD): This is the sum of DA and DE (plus a negligible term that disappears as sample size increases). It represents the overall difference between the population covariance matrix ($\mathbf{\Sigma}_0$) and the model's implied covariance matrix in the sample ($\hat{\mathbf{\Sigma}}$). OD takes into account both the model's lack of fit in the population (DA) and the sampling variability (DE). Minimizing OD is equivalent to maximizing the model's generalizability or its ability to fit new data arising from the same population. It quantifies the overall discrepancy between the population covariance matrix and the model-implied covariance matrix estimated from a sample.

At small sample sizes, DA and DE can have opposite effects on OD. Adding parameters may reduce DA but increase DE. As sample size increases, DE decreases, and OD approaches DA. Therefore, minimizing DA and OD are asymptotically equivalent, but at realistic sample sizes, the model with the lowest DA may not have the lowest OD.

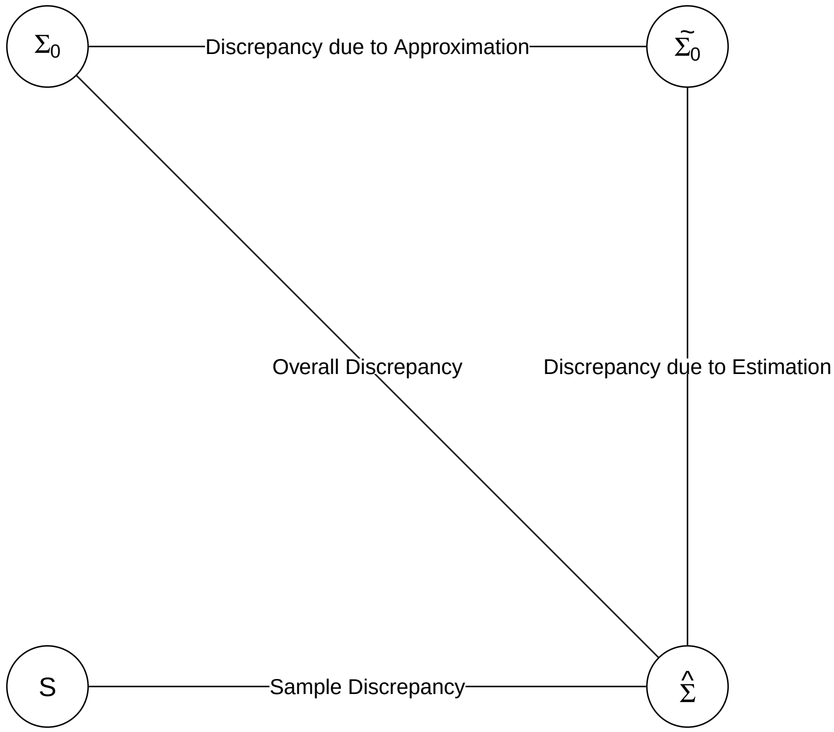


Figure 2

Different types of discrepancy according to Cudeck and Henly (1991).

Greene et al. (2019), Morgan et al. (2015), and Bonifay and Cai (2017) showed that traditional fit indices are biased in favor of bifactor models. Similarly, Kan et al. (2024) argued that approximate fit indices (like CFI and RMSEA), which reflect sample discrepancy between the sample covariance matrix and the model-implied covariance matrix, should not be used for comparing models. Instead, Kan et al. (2024) propose that relative fit metrics like AIC and BIC, which approximate overall discrepancy between the population covariance matrix ($\boldsymbol{\Sigma}_0$) and $\boldsymbol{\Sigma}$, are more accurate in differentiating bifactor models from other types of models. However, these studies share a common limitation: they focus solely on the accuracy of fit indices in selecting the model with the highest generalizability (i.e., minimizing OD), failing to consider the equally important goal of selecting the model with the highest verisimilitude (i.e., minimizing DA). As Cudeck and Henly (1991) and Preacher et al. (2013) argue, these two goals are distinct and may lead to different model choices, especially when different sample sizes are investigated. Therefore, a comprehensive evaluation of fit indices in the context of bifactor modeling should consider their performance in relation to both generalizability and verisimilitude.

When the focus is on one type of evidence because the data generation mechanism can capture only model divergence due to estimation (i.e., the data is generated from a specific model-implied covariance matrix that is a perfect representation of the population covariance matrix), nothing can be said about the fit indices' capacity to identify divergence due to approximation. Divergence due to approximation is only possible when the data generation mechanism has a model-implied covariance matrix that is not a perfect representation of the population covariance matrix. In other words, it is necessary to add error to the population to be able to capture more realistic evidence about divergence due to estimation (sample-level) and to capture divergence due to approximation (population level). Greene et al. (2019), Morgan et al. (2015), Bonifay and Cai (2017), and Kan et al. (2024) did not include error in the population, nor did they investigate how different fit indices perform at the population level.

Present Study

Golino et al. (2021) recently developed a new family of fit indices for dimensionality analysis called entropy fit indices. These indices, based on information and quantum information theory, provide an objective function that can be optimized for dimensionality tasks. In their simulation study, the total entropy fit index (*TEFI*) demonstrated the highest power in detecting the correct (population) structure compared to other entropy fit indices and traditional fit measures used in structural equation modeling, such as the CFI and RMSEA.

TEFI computes the distance between the mean Von Neumann (*VN*) entropy for the estimated dimensions (factors) and the total *VN* entropy of the system of variables, with a penalization for the number of dimensions. This penalization ensures that *TEFI* only decreases with additional dimensions if each dimension reduces the uncertainty or disorder of the system of variables. *TEFI* requires only a vector representing the structure (i.e., how variables are grouped into dimensions) and a scaled correlation matrix. It measures the entropy reduction by partitioning a multidimensional space into several groups of variables, aiming to minimize the uncertainty or disorder due to the correct identification of underlying dimensions (Golino et al., 2021).

Although *TEFI* overcomes some limitations of traditional fit indices, it is only suitable for single-level structures, such as those estimated using exploratory graph analysis (Christensen & Golino, 2021; EGA; Golino & Epskamp, 2017; Golino et al., 2020), reflecting a correlated traits structure. In this paper, we introduce the generalized total entropy fit index (*GenTEFI*), which expands *TEFI* to accommodate more complex two-level structures represented by a bifactor organization with multiple general factors, such as those estimated using hierarchical EGA (Jiménez, Abad, Garcia-Garzon, Golino, et al., 2023).

Our simulation studies address the limitations of recent simulations by generating data from a population model with error, enabling the quantification of fit indices' accuracy in capturing discrepancy due to sampling and approximation. We check the accuracy of fit indices in three simulations:

1. True data generation mechanism: correlated traits model with population error.
2. True data generation mechanism: bifactor model with population error.
3. True data generation mechanism: bifactor model with population error, comparing the correct structure to over- or under-factored structures.

In all simulations, we compute the accuracy of fit indices at both the sample and population levels. Trustworthy fit indices should demonstrate high accuracy at both levels, indicating a high capacity to select models with generalizability and verisimilitude. Fit indices that perform well in samples but poorly in the population (or vice versa) are biased and should be used with caution.

We show that *GenTEFI* is the only fit index that can adequately differentiate between correlated traits and bifactor structures when the data generation mechanism is either a correlated traits model or a bifactor model, both at the sample and population levels. Additionally, when the data generation mechanism is a bifactor model, *GenTEFI* performs as well as traditional fit indices but with higher accuracy in identifying the correct structure when the general factor is overfactored.

The paper is organized as follows. First, we present *TEFI* and its expansion, *GenTEFI*. Then, we discuss traditional fit indices used in dimensionality assessment and briefly introduce bifactor models with multiple correlated general factors. We describe three Monte Carlo simulations, present the results, and provide an empirical example. Finally, we discuss the findings, limitations, and future directions.

The Total Entropy Fit Index and its generalization to two-level structures

Entropy, Total Correlation, and the K-Function

Golino et al. (2021) introduced a new a family of fit indices based on the information-theoretic concept of entropy, that can be used to check the fit of the dimensionality structures estimated via EGA (Christensen & Golino, 2021; Golino & Epskamp, 2017; Golino et al., 2020) and factor-analytic methods. In information theory, entropy is a measure of the uncertainty associated with a random variable. It is a way of quantifying the amount of randomness or unpredictability in a system, with its historical roots in thermodynamics with the work of Boltzmann and others (Watanabe, 1960).

The entropy of a discrete random variable X depicts the average amount of unpredictability that is removed when the outcome of X is known (Yeung, 2008), characterizing the *uncertainty* associated with X or the degree of disorganization of X. Watanabe (1939) proposed an entropy-based index termed *total correlation* to measure the uncertainty of random variables and the strength of their correlation beyond their average interaction (Watanabe, 1960). The total correlation of a set of variables increases as the entropy of individual variables increases or as the entropy of the set of variables decreases (Watanabe, 1960). Replacing metrics of distance by total correlation, Watanabe (Watanabe, 1969) developed a cluster-identification technique termed *interdependence analysis.* The general idea of interdependence analysis is to partition a set into subsets until interdependence is minimized (Watanabe, 2000).

Hiroshi Watanabe (Watanabe, 2000) showed that constructing clusters using total correlation could lead to undesirable outcomes in partitioning a multidimensional space of variables. As total correlation disregards the size of the subsets, small subsets with low entropy values might be removed, regardless of their relationship to other subsets. To address this issue, Hiroshi Watanabe (Watanabe, 2000, 2001) introduced the *K-function*, a modified total correlation index calculated as the difference between the average entropy of X_v and X_ω from the entropy of the super-set $\mathcal{X}$, $(X_v, X_\omega) \in \mathcal{X}$.

The *K-function* can be used as a measure of entropy reduction by partitioning, where minimization reduces the uncertainty or instability in the partition of a multidimensional space, and can also be interpreted as the maximization of information gain (Watanabe, 2000, 2001). However, despite its interesting properties, Golino et al. (2021) showed that the *K-function* is not suitable as a fit index for dimensionality assessment because it decreases as the true number of subgroups of variables (i.e., dimensions, clusters, factors, or communities) increases.

Total Entropy Fit Index

Golino et al. (2021) developed a family of fit indices to overcome the limitations of the *K-function* proposed by Watanabe (2000) and Watanabe (2001). One of these new fit indices, the total entropy fit index (*TEFI*), presented the highest power in a Monte Carlo simulation. When *TEFI* was used to compare the correct (population) structure with incorrect structures (with more or less factors than the population model used to generate the data, or with a different composition of items per factor), it presented a mean percentage correct score (i.e., the mean number of times it selected the correct dimensionality structure when compared to incorrect structures) of 92%. RMSEA and CFI presented a mean percentage correct score of 78% and 74% when used as relative measures of fit, and 14% and 35% when their traditional cut-off criteria ($\leq .05$ and $\geq .95$, respectively) was used.

The *TEFI* index computes the distance between the mean Von Neumann entropy (Von Neumann, 1927) of the dimensions (factors or communities) and the total entropy of the system of variables, and adds a penalization to the number of dimensions estimated. Von Neumann entropy (Von Neumann, 1927) is 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. As Golino et al. (2021) points out, 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 $\boldsymbol{\rho}$, Von Neumann entropy is:

$$\mathcal{S}(\boldsymbol{\rho}) = -tr(\boldsymbol{\rho}\log\boldsymbol{\rho}). \tag{1}$$

Golino et al. (2021) shows that 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, and can be used to describe or quantify properties of a system of variables (Golino et al., 2021).

Wihler, Bessire, and Stefanov (2014) shows that Von Neumann entropy can also be estimated using the eigenvalues of a density matrix. Given a density matrix $\boldsymbol{\rho}$ with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_m \geq 0$, Von Neumann entropy is:

$$\mathcal{S}(\boldsymbol{\rho}) = -tr(\mathcal{L}(\mathbf{D})). \tag{2}$$

in which $\mathbf{D}$ is a density matrix, and $\mathcal{L}$ is the vector eigenvalues of the density matrix. In other words, the Von Neumann entropy of a density matrix is the Shannon entropy of the vector of eigenvalues (Golino et al., 2021; Preskill, 2018). Additionally, the joint Von Neumann entropy of two (or more) density matrices, $\mathbf{A}, \mathbf{B}$, can be calculated as products of their individual eigenvalues, since the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are $\lambda_1\mu_1, \cdots, \lambda_1\mu_m, \cdots, \lambda_2\mu_1, \cdots, \lambda_2\mu_m, \cdots, \lambda_n\mu_m$ (Laub, 2005).

The *TEFI* index can be computed using a density-like, scaled version of a correlation matrix in which the Von Neumann entropy can be estimated on. Given a $m \times m$ scaled correlation matrix $\mathbf{p}$ with two underlying dimensions, factors, or communities indexed by η_1 and η_2, these dimensions can be represented as a different combination of sub-elements (rows and columns) from $\mathbf{p}$: $\mathbf{p}_{i,j\in\eta_1}$ and $\mathbf{p}_{i,j\in\eta_2}$, with $\eta_1 = (1, 2, 3)$, and $\eta_2 = (4, 5, 6)$. In other words, the first dimension (factor or community) is composed by variables 1, 2 and 3 of $\mathbf{p}$, and the second factor is composed by variables 4, 5 and 6 of $\mathbf{p}$. With $m = 6$, $\mathbf{p}$ takes the generic form of

$$\mathbf{p} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{bmatrix}.$$

The sub-matrix $\mathbf{p}_{i,j\in\eta_1}$ is, therefore:

$$\mathbf{p}_{i,j\in\eta_1} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

And the sub-matrix $\mathbf{p}_{i,j\in\eta_2}$ is:

$$\mathbf{p}_{i,j\in\eta_2} = \begin{bmatrix} a_{44} & a_{45} & a_{46} \\ a_{54} & a_{55} & a_{56} \\ a_{64} & a_{65} & a_{66} \end{bmatrix}.$$

The Von Neumann entropy of $\mathbf{p}_{i,j\in\eta_1}$ and $\mathbf{p}_{i,j\in\eta_2}$ are $\mathcal{S}(\mathbf{p}_{\eta_1})$ and $\mathcal{S}(\mathbf{p}_{\eta_2})$, respectively (with $i,j \in \eta_1$ and $i,j \in \eta_2$ omitted for clarity). Then, *TEFI* can be calculated as:

$$TEFI = \left[\frac{\sum_{k=1}^{N_F} \mathcal{S}(\mathbf{p}_k)}{N_F} - \mathcal{S}(\mathbf{p})\right] + \left[\left(\mathcal{S}(\mathbf{p}) - \sum_{k=1}^{N_F} \mathcal{S}(\mathbf{p}_k)\right) \times \sqrt{N_F}\right], \tag{3}$$

where $\mathcal{S}(\mathbf{p})$ is the total Von Neumann entropy of the system of variables (i.e., the Von Neumann entropy of the scaled correlation matrix with all variables), $\mathcal{S}(\mathbf{p}_k)$ is the Von Neumann entropy of the scaled sub-matrix representing dimension (factor or community) k, and N_F is the number of dimensions (or factors, communities).

This equation integrates three key components: the average entropy across dimensions or factors (A), the number of dimensions (B), and the total entropy of the system (C). The total entropy fit index, can be re-written as $TEFI = [P1] + [P2]$, in which $[P1] = \frac{A}{B-C}$ and $[P2] = (C - A) \times \sqrt{B}$. Golino et al. (2021) argues that $[P1]$ is expected to decrease monotonically as the number of factors increases, but $[P2]$ is expected to increase as the number of factors increase, representing the reduction in average entropy of a set of data conditional on a given dimensional (factor) structure. Finally, the square root of the number of factors was chosen in $[P2]$ by Golino et al. (2021) to control the expected growth trajectory of $[P2]$ as the number of factors increases. 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. In summary, TEFI computes the distance between the mean Von Neumann entropy of the dimensions (factors or communities) and the total entropy of the system of variables, and adds a penalization to the number of dimensions estimated.

Generalized Total Entropy Fit Index

As can be noted by equation 3, TEFI can only be computed in single-level structures (e.g., those reflecting first-order factors or communities). At present, there's no way for TEFI to take into consideration more complex dimensionality structures, such as bifactor structures with multiple correlated general factors. Expanding TEFI to accommodate dimensionality structures in two levels requires adding two new components to equation 3. One, $[P3]$, representing the distance (or difference) between the sum of the Von Neumann entropy of the *second-level* dimensions, E, relative to (i.e., divided by) the number of *first-level* dimensions (first-order factors, group factors, first-order communities), B, and the total entropy of the system of variables (C). The other component to be added ($[P4]$), is similar to the penalization component $[P2]$ of TEFI, but replaces A by E (the sum of the individual Von Neumann entropy of the second-level dimensions). The simplified version of

the generalized total entropy fit index is, therefore,
$GenTEFI = [\left(\frac{A}{B-C}\right) + (C - A) \times \sqrt{B}] + [\left(\frac{E}{B-C}\right) + (C - E) \times \sqrt{B}]$, or
$GenTEFI = [P1] + [P2] + [P3] + [P4]$, being $[P3] = (E/B) - C$ and $[P4] = (C - E) \times \sqrt{B}$.
The generalized total entropy fit index can be seen as an additive fit index combining a first-order TEFI ($TEFI_{First-Order} = [P1] + [P2]$) and a second-order TEFI ($TEFI_{Second-Order} = [P3] + [P4]$).

The generalized TEFI can be, finally, formulated as:

$$\begin{aligned} GenTEFI = &\left\{ \left[\frac{\sum_{k=1}^{N_F} \mathcal{S}(\mathbf{p}_k)}{N_F} - \mathcal{S}(\mathbf{p}) \right] + \left[\left(\mathcal{S}(\mathbf{p}) - \sum_{k=1}^{N_F} \mathcal{S}(\mathbf{p}_k) \right) \times \sqrt{N_F} \right] \right\} \\ &+ \left\{ \left[\frac{\sum_{l=1}^{N_{GF}} \mathcal{S}(\mathbf{p}_l)}{N_F} - \mathcal{S}(\mathbf{p}) \right] + \left[\left(\mathcal{S}(\mathbf{p}) - \sum_{l=1}^{N_{GF}} \mathcal{S}(\mathbf{p}_l) \right) \times \sqrt{N_F} \right] \right\}, \end{aligned} \tag{4}$$

where N_{GF} is the number of second-level dimensions and $\mathcal{S}(\mathbf{p}_l)$ is the sum of the Von Neumann entropies for each factor or community l of the second level structure. Equation 4 can be simplified as follows:

$$GenTEFI = \left[\frac{\sum_{k=1}^{N_F} \mathcal{S}(\mathbf{p}_k) + \sum_{l=1}^{N_{GF}} \mathcal{S}(\mathbf{p}_l)}{N_F} - 2 \times \mathcal{S}(\mathbf{p}) \right] + \left[\left(2 \times \mathcal{S}(\mathbf{p}) - \sum_{k=1}^{N_F} \mathcal{S}(\mathbf{p}_k) - \sum_{l=1}^{N_{GF}} \mathcal{S}(\mathbf{p}_l) \right) \times \sqrt{N_F} \right]. \tag{5}$$

Interesting Properties of The Generalized Total Entropy Fit Index

To demonstrate the applicability of the GenTEFI in comparing structures with varied organizations—such as correlated traits versus bifactor models—data were generated under two distinct conditions. In the first scenario, we generated data using a correlated traits structure, characterized by factor loadings between 0.45 and 0.75, correlations between factors ranging from 0.00 (orthogonal) to 0.70, and four variables per factor in a total of four factors. In the second scenario, the data was generated using a bifactor model comprising four group factors and one general factor, with four variables per group factor. Here, the loadings on the group factors varied from 0.45 to 0.55, while those on the general factors ranged from 0.45 to 0.70. For each model, we generated 100 datasets, each consisting of 5,000 observations, to robustly assess the characteristics of the GenTEFI (more specifically, the lower-order or first-order TEFI and the high-order or second-order TEFI values) across these different structural configurations.

In Figure 3, a hexagonal binning plot shows the relationship between factor loadings and the first and second-order TEFI under varying conditions of a correlated traits model. The use of hexagonal cells enables a clear depiction of data point density, with color gradients indicating the concentration of points—brighter yellow hues signifying greater densities. The plot's facets correspond to different levels of interfactor correlation, ranging from orthogonal (zero correlation) to highly correlated factors, enabling a comparative analysis across the spectrum of loading magnitudes. The hexagons' borders are color-coded to denote the level of TEFI: red for second-order TEFI and gray for first-order TEFI ($TEFI_{First-Order} = [P1] + [P2]$, $TEFI_{Second-Order} = [P3] + [P4]$, see equation 4).

The figure reveals that, for orthogonal or weakly correlated trait structures, the first-order TEFI values are consistently lower than those of the second-order meaning the *uncertainty* of the correlated traits structure is *lower* than the bifactor structure. The first-order TEFI was computed in a structure that mirrors the true four-factor model whereas the second-order TEFI is based on an assumed but incorrect higher-order factor. Interestingly, as interfactor correlations strengthen, the gap between first and second-order TEFI narrows. At higher interfactor correlations, the second-order TEFI becomes less than the first-order, suggesting an emergent second-order structure. These observations underscore two key insights regarding the GenTEFI: its sensitivity to distinguishing between correlated traits and bifactor or hierarchical structures and its responsiveness to changes in the correlation between factors, which reflects different organizational complexities within the data structure. As the correlation increases, more information is being shared between the factors, which makes them more mixed or presents a higher disorganization and higher uncertainty. As in physical systems, systems of variables with mixed states show a higher entropy, disorganization, or uncertainty than systems well-compartmentalized into distinct sets. Consequently, with increased correlations between factors, the overall system of variables has a higher entropy or uncertainty than the average entropy of its individual factors. This suggests that the global interactions of the variables contribute more to its entropy than the internal structure of its factors. Consequently, the uncertainty of a hypothetical second-order factor decreases. In other words, as the correlations between factors increase, a second-order structure emerges.

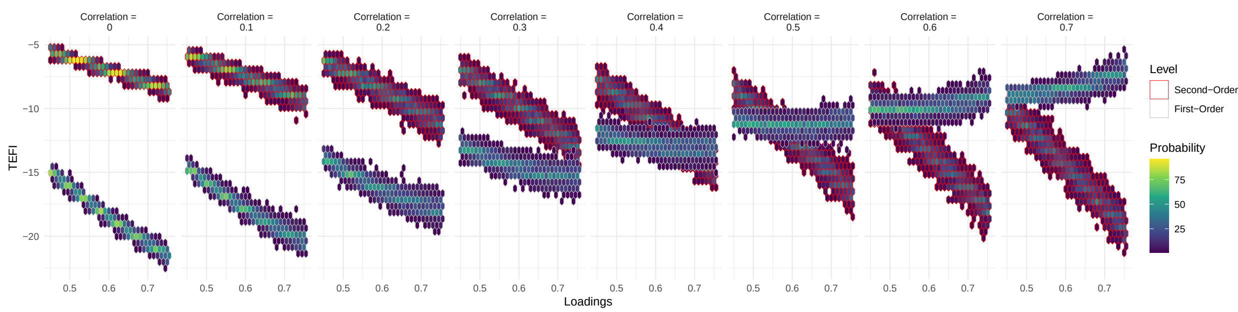


Figure 3

Total entropy fit values accross different conditions used to generate data from a correlated traits model per level (first-order and second-order).

In Figure 4, a hexagonal binning plot shows the relationship between factor loadings and the Total Entropy Fit Index (`TEFI`) under varying conditions of a bifactor model. Loadings of the general factors are represented in the x-axis, and the grid represent different magnitude of the first-order factor loadings (or loadings of the group factors). The plot indicates that the values for second-order TEFI are lower than for the first-order ones, and this gap grows as the general factors' loadings increase. This distinction between the first and second-order TEFI values highlights the usefulness of the `GenTEFI`. In short, both Figures (3 and 4) illustrate that `GenTEFI` is an effective new tool for analyzing the dimensionality of psychology data—it helps differentiate between a correlated traits model and a bifactor model. When the underlying structure of the data is a correlated traits model, first-order TEFI values are smaller (more negative) than second-order ones. For example, using GenTEFI on the structures depicted in Figure 1 (correlated traits and the bifactor structure), leads to a lower-order TEFI of -13.03 and a high-order TEFI of -8.03, indicating that the correlated traits structure fits the data better than the bifactor structure (or that the former has a lower uncertainty than the latter), contrary to the traditional fit indices shown earlier in the introduction. At the same time, when dealing with data generated from a bifactor model, GenTEFI proves useful in assessing the fit of both levels (first and second order) simultaneously, as the simulation studies in the last three sections of the paper will demonstrate.

Figure 4

Total entropy fit values accross different conditions used to generate data from bifactor model per level (first-order and second-order)

Fit Indices used in Factor Analysis

Fit indices developed under the factor analysis framework have been used as tools of dimensionality assessment (Santos, Vagos, & Rijo, 2018; Schermelleh-Engel, Moosbrugger, & Müller, 2003; Ventimiglia & MacDonald, 2012). These fit indices include, among others: CFI (Bentler, 1990), RMSEA (Steiger, 1980), and SRMR (Joreskog & Sorbom, 1981). CFI tests the equivalence of the population covariance matrix of observed variables and the model-implied covariance matrix using the noncentrality parameter (λ_M) of the specified model, which is computed as $\chi^2_M - df_M$, where χ^2_M is the chi-square statistic and df_M is the degrees of freedom of the specified model. SRMR computes the average difference between observed and model-implied covariance matrices. RMSEA and SRMR are considered absolute fit indices because they compute an overall fit that only takes into account the fit of the specified model. CFI is an incremental fit index because it verifies the superiority of the specified model relative to a baseline model at reproducing the observed covariances.

The formulas for these fit indices are shown below in Equations 6, 7, and 8,

$$CFI = 1 - \frac{max(\lambda_M, 0)}{max(\lambda_N, \lambda_M)}, \tag{6}$$

$$\text{RMSEA} = \max\left(\sqrt{\frac{\lambda_M}{df_M(N-1)}}, 0\right), \tag{7}$$

$$SRMR = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}\left(\frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} - \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}}\sqrt{\hat{\sigma}_{jj}}}\right)^2}{\frac{p(p+1)}{2}}}, \tag{8}$$

where λ_M and df_M are the noncentrality parameter and degrees of freedom of the specified model, 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.

Applied researchers (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 .05 or .08 for RMSEA and SRMR (good and acceptable fit, respectively; Browne & Cudeck, 1992; Chen, Curran, Bollen, Kirby, & Paxton, 2008; Hu & Bentler, 1999) to verify the dimensionality structure of their instruments. However, recent simulation studies have shown that using cut-off values as criteria for dimensionality assessment is not adequate (Garrido, Abad, & Ponsoda, 2016; Golino et al., 2021). Clark and Bowles (2018) and Beierl, Bühner, and Heene (2018), for example, showed that none of the conventional cut-off values previously cited appeared to perform well enough to be recommended, a finding replicated by Golino et al. (2021). Clark and Bowles (2018), Beierl et al. (2018), and Garrido et al. (2016) 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, although CFI performed more reliably using a 0.95 cut-off value, its 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.

Golino et al. (2021) found that despite not performing well when using the cut-off values previously cited, CFI and RMSEA performed very well when used as relative measures of fit. The use of traditional fit indices as relative measures of fit means that two or more dimensionality structures are being compared, and the one presenting the highest CFI value, or the lowest RMSEA/SRMR fits the data better. The values of the fit indices are affected by 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; Golino et al., 2021; Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011; Hu & Bentler, 1999; McNeish, An, & Hancock, 2018; Savalei, 2012; Xia & Yang, 2018).

Golino et al. (2021), however, did not investigate the accuracy of relative measures of fit, such as Akaike's information criterion (AIC; Akaike, 1973) and the Bayesian information criterion (BIC; Schwarz, 1978). Both fit indices are part of a broader set of fit indices based on overall discrepancy (OD), designed to select the simplest model from a pool of competing models that most describes observed data more accurately (Preacher et al., 2013), and can be generally represented with the following equation:

$$OD = -2f_k + \alpha \times q_k, \tag{9}$$

where f_k is the log-likelihood of model k, q_k is the number of free parameters in model

k, and α is a function of sample size: in AIC $\alpha = 2$ and in BIC $\alpha = ln(N)$, N being sample size.

AIC and BIC are two model selection criteria that impose different complexity penalties, with BIC having a stiffer penalty than AIC for sample sizes greater than or equal to 8 (Preacher et al., 2013). Although AIC has been shown to perform well in selecting the true number of factors at small sample sizes, it tends to select more complex models as the sample size increases (Song & Belin, 2008). This tendency has raised questions about the appropriateness of AIC for model selection, as it was designed with generalizability in mind (ability to cross-validate well to data arising from the same underlying process) rather than verisimilitude (proximity to the objective truth; Preacher et al., 2013). BIC, on the other hand, has been found to outperform AIC in recovering the true number of factors, but suffered as N increased to large sample sizes (Preacher et al., 2013).

To our knowledge, the current paper is the first to evaluate how well traditional fit indices used in factor analysis (CFI, RMSEA, SRMR, AIC, and BIC) perform as tools of dimensionality assessment in bifactor structures, computing accuracy at the sample and population levels, and using population covariance matrices with error. Therefore, there are no previous Monte Carlo simulations to base our expectations on or compare our results to. Nonetheless, we expect to replicate—to some extent—the results of the previous simulations using the more common correlated first-order factor structures.

Bifactor models with multiple general factors

Bifactor analysis (BIFA) is a popular approach to decompose item variance in group factors that affect a subset of items and one orthogonal general factor that is common to all the items (Reise, 2012). This modeling framework has been used in a wide variety of fields, like personality (Abad, Sorrel, Garcia, & Aluja, 2018) and intelligence (Cucina & Byle, 2017b; Eid, Krumm, Koch, & Schulze, 2018), to deal with the simultaneous presence of narrow and broad dimensions, giving a comprehensive view of the psychological traits that are measured.

However, BIFA is restricted to a single general factor, whereas psychometric instruments are often created with items that encompass many different domains. For this reason, with the prolific development of BIFA methods in the last decade (Abad, Garcia-Garzon, Garrido, & Barrada, 2017a; Eid, Geiser, Koch, & Heene, 2017; Garcia-Garzon, Abad, & Garrido, 2019; Garcia-Garzon, Nieto, Garrido, & Abad, 2020; Jennrich & Bentler, 2011, 2012; Lorenzo-Seva & Ferrando, 2019; Waller, 2018), there has also been an increasing interest to address more than one general factor in a single model (Cai, 2010; Jiménez, Abad, Garcia-Garzon, & Garrido, 2023a; Tian & Liu, 2021). In the context of bi-factor analysis with multiple general factors (BIFA-MGF: Jiménez, Abad, Garcia-Garzon, & Garrido, 2023b), a general factor is defined as a dimension whose indicators are themselves indicators of at least two group factors, giving a nested-looking structure in which the variance shared by the group factors is contained in the general factors. This generalization of BIFA takes into account the dependencies between the domains (i.e., correlations between the general factors and interstitial cross-loadings), removing possible biases that may appear when conducting independent BIFA analyses for each domain of the data. Furthermore, BIFA-MGF can be used in a completely exploratory framework (Jiménez, Abad, Garcia-Garzon, & Garrido, 2023a), facilitating the discovery of theoretical misspecifications (e.g., items loading on a different general factor than expected by the theory) that will promote the content expansion or redefinition of broad traits. Despite the appealing features of this bi-factor extension, estimating these models require the correct specification of the number of group and general factors, which can be challenging in many situations (but see Jiménez, Abad, Garcia-Garzon, Golino, et al., 2023).

Methods

Simulation 1: Correlated Traits (true data generation mechanism) vs. Bifactor Structures

To investigate the comparative performance of the GenTEFI with respect to the other fit indices (CFI, RMSEA, SRMR, BIC, AIC) in the differentiation of correlated traits and bifactor structures, a simulation was implemented having a correlated traits (with four factors) as the data generation mechanism. We manipulated four variables to create a fully crossed design with 24 conditions: (a) Number of variables per factor (`N.VAR`: 4, 8), (b) loadings ("low", ranging from 0.35 to 0.45, "moderate", ranging from 0.45 to 0.55, and "high", ranging from 0.6 to 0.7), (d) correlations between factors (`Cor`: 0, 0.30, 0.50), and (e) sample size (`N`: 500, 1000, 5000). For each combination of conditions, 500 datasets were generated with population error following the method of Cudeck and Browne (1992) using the `sim_factor` function from the `bifactor` package (version 0.1.0; Jimenez, Abad, Garcia-Garzon, Garrido, & Franco, 2022).

For each condition, two structures were estimated, one reflecting the correct four-factor structure used to generate the data, and one reflecting a bifactor structure with four first-order orthogonal factors and one general factor. The traditional fit indices (CFI, RMSEA, SRMR, BIC, AIC) were computed using confirmatory factor analysis via `lavaan` (version 0.6.18; Rosseel, 2012a). The fit indices (CFI, RMSEA, SRMR, BIC, AIC, and GenTEFI) were computed for each structure (correlated traits and bifactor) and when the indices indicated that the correct (correlated traits) structure fitted the data better than the incorrect structure (bifactor), an accuracy score of one was assigned (otherwise, the accuracy was zero, indicating the selection of the incorrect structure). The simulated data, for this simulation, was continuous.

Simulation 2: Bifactor Structures (true data generation mechanism) vs. Correlated Traits

The comparative performance of the GenTEFI with respect to the other fit indices (CFI, RMSEA, SRMR, BIC, AIC) in the differentiation of bifactor and correlated trait structures, a simulation was implemented having a bifactor model (with four group factors and one general factor) as the data generation mechanism. We manipulated four variables to create a fully crossed design with 40 conditions: (a) Number of variables per factor (`N.VAR`: 4, 8), (b) loadings in the group factors (`LOADS.GRF`: "low", ranging from 0.35 to 0.45, and "moderate", ranging from 0.45 to 0.55), (c) loadings in the general factor (`LOADS.GF`) centered in 0.45 (ranging from 0.35 to 0.55), 0.50 (ranging from 0.40 to 0.60), 0.55 (ranging from 0.45 to 0.65), 0.60 (ranging from 0.50 to 0.70), 0.65 (ranging from 0.55 to 0.75), and (d) sample size (`N`: 500, 1000, 5000). Generating the loadings randomly from distributions helped break the proportionality constraint of higher-order models and ensured that the simulated data were full rank bifactor structures (Gignac, 2016). Additionally, each factor contained one approximately pure indicator of the general factor to further break the proportionality constraint (Abad, Garcia-Garzon, Garrido, & Barrada, 2017b). This item had a negligible loading of 0.10 on its assigned group factor. For each combination of conditions, 500 datasets were generated with population error following the method of Cudeck and Browne (1992) using the `sim_factor` function from the `bifactor` package.

For each condition, two structures were estimated, one reflecting the correct bifactor structure used to generate the data, and one reflecting a correlated trait structure with four first-order factors. The traditional fit indices (CFI, RMSEA, SRMR, BIC, AIC) were computed using confirmatory factor analysis via `lavaan`. The fit indices (CFI, RMSEA, SRMR, BIC, AIC, and GenTEFI) were computed for each structure (correlated traits and bifactor) and when the indices indicated that the correct (bifactor) structure fitted the data better than the incorrect structure (correlated traits), an accuracy score of one was assigned (otherwise, the accuracy was zero, indicating the selection of the incorrect structure). The simulated data, for this simulation, was continuous.

Simulation 3: Bifactor Structures vs. Bifactor Structures

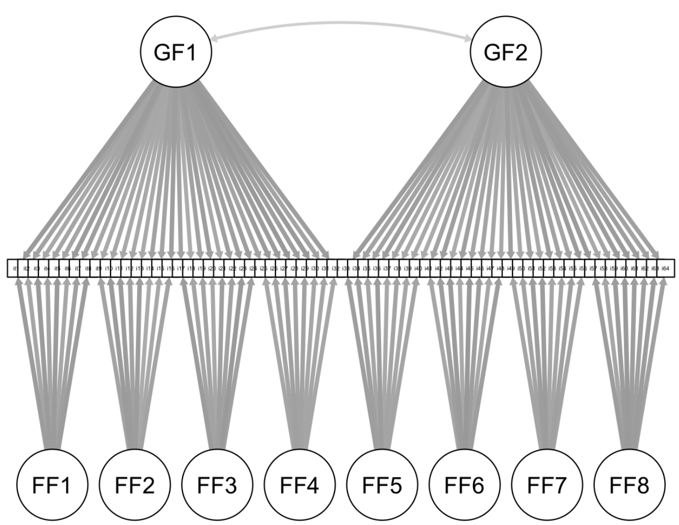
To investigate the comparative performance of the GenTEFI with respect to the other fit indices (CFI, RMSEA, SRMR, BIC, AIC), we manipulated five variables to create a fully crossed design with 48 conditions: (a) Number of variables per group factor (`VAR.GRF`: 4, 8), (b) loadings on the general factors (`LOAD.GF`: "low", "medium"), (c) loadings on the group factors (`LOAD.GRF`: "low", "medium"), (d) correlations between the general factors (`COR.GF`: 0.30, 0.50, 0.70), and (e) sample size (`N`: 500, 1000, 5000).The factor loadings were generated using a uniform distribution centered in 0.40 (from 0.35 to 0.45) for the low condition and centered in 0.50 (from 0.45 to 0.55) for the medium condition in the group factors. In the general factors, the low loading condition was generated using a uniform distribution ranging from 0.30 to 0.50, and the moderate condition from 0.50 to 0.70. Generating the loadings randomly from distributions helped break the proportionality constraint of higher-order models and ensured that the simulated data were full rank bifactor structures (Gignac, 2016). Additionally, each factor contained one approximately pure indicator of the general factor to further break the proportionality constraint (Abad et al., 2017b). This item had a negligible loading of 0.10 on its assigned group factor. For each combination of conditions, 500 datasets were generated with population error following the method of Cudeck and Browne (1992) using the `sim_factor` function from the `bifactor` package.

The traditional fit indices (CFI, RMSEA, SRMR, BIC, AIC) were computed using confirmatory factor analysis via `lavaan`. All the generated bifactor structures had two general factors and eight group factors, with the items pertaining to four group factors being also indicators of one general factor (see Table 1 and Figure 5).

Table 1

A random sample of simulated medium-sized loadings for a bi-factor structure with two general factors and eight group factors.

Item	First Domain					Item	Second Domain				
	G1	S1	S2	S3	S4		G2	S5	S6	S7	S8
1	.51	.10	0	0	0	17	.60	.10	0	0	0
2	.50	.52	0	0	0	18	.59	.52	0	0	0
3	.50	.54	0	0	0	19	.51	.53	0	0	0
4	.51	.53	0	0	0	20	.51	.58	0	0	0
5	.53	0	.58	0	0	21	.54	0	.53	0	0
6	.52	0	.52	0	0	22	.60	0	.10	0	0
7	.58	0	.53	0	0	23	.52	0	.56	0	0
8	.59	0	.10	0	0	24	.57	0	.53	0	0
9	.56	0	0	.55	0	25	.57	0	0	.57	0
10	.60	0	0	.10	0	26	.59	0	0	.57	0
11	.54	0	0	.59	0	27	.60	0	0	.10	0
12	.53	0	0	.56	0	28	.57	0	0	.59	0
13	.51	0	0	0	.56	29	.54	0	0	0	.55
14	.51	0	0	0	.59	30	.56	0	0	0	.10
15	.52	0	0	0	.10	31	.52	0	0	0	.53
16	.50	0	0	0	.58	32	.56	0	0	0	.60



Population Error

To simulate more realistic data, the model correlation matrix of each condition was contaminated with error, creating some misspecification. Model misspecification is ubiquitous in psychological modeling. The current consensus admits that it is impossible to accommodate the presence of the many nuances and minor factors that explain common item variance beyond the major factors that are specified in practice (MacCallum, 2003; Mõttus et al., 2020). At best, factor models can capture the broad traits underlying the data. According to this perspective, the population correlation matrix can be expressed as

$$\mathbf{R} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^{\top} + \mathbf{\Psi} + \mathbf{E}, \tag{10}$$

where $\mathbf{\Lambda}$ is the item loading matrix on the major factors, $\mathbf{\Phi}$ is the matrix of major factors correlations, $\mathbf{\Psi}$ is the diagonal matrix with unique item variances, and $\mathbf{E}$ is a symmetric, off-diagonal matrix of covariances for which the minor factors are responsible (i.e, an error term).

A consequence of the inevitable misspecification carried in the $\mathbf{E}$ term is that the model parameter estimates, even when they are unbiased, cannot reproduce the population correlation matrix. Therefore, to face this situation, applied researches usually evaluate the difference between the population and model correlation matrices with fit indices to decide whether the amount of misspecification is small enough as to render the model an acceptable or good simplification of the reality (Maydeu-Olivares, Shi, & Rosseel, 2018; Shi, Maydeu-Olivares, & Rosseel, 2020).

Such model assessment with fit indices has motivated the development of several methods intended to generate data (i.e., the $\mathbf{E}$ term) with varying levels of misspecification according to different values of the fit indices, so that the datasets created in simulation research resemble better the messy data encountered in practice (Cudeck & Browne, 1992;

Lai, 2019; Tucker, Koopman, & Linn, 1969; Yuan & Hayashi, 2003). In fact, many studies have already uncovered the fallibility of many dimensionality methods under the presence of model misspecification (Jiménez, Abad, Garcia-Garzon, Golino, et al., 2023; Lim & Jahng, 2019; Xia, 2021), strengthening the view that simulation studies in dimensionality research should include a condition involving population error to avoid erroneous conclusions.

For this simulation, we decided to generate data with population error following the method of Cudeck and Browne (1992) using the `sim_factor` function from the `bifactor` package. This approach has two features that make it more appealing than other competing methods. First, the objective value of the discrepancy function or a given fit index can be fixed, giving a precise control of the desired amount of misspecification. Second, the global minimum at the original model parameters is preserved after adding the $\mathbf{E}$ term, so that the nominal conditions involving the magnitude of item loadings, cross-loadings, and factor correlations are not contaminated.

The amount of population error was specified according to the CFI index estimated via maximum likelihood, with levels of CFI = 0.95 chosen to reflect an excellent fit to the data, and CFI = 0.90 to represent a good level of fit (Montoya & Edwards, 2021). Even though all commonly used factor-analytic fit indices are impacted by incidental parameters not related to the size of the misfit, the CFI has proven to be the most robust to latent misspecification in comparison to absolute fit indices such as RMSEA or SRMR (Garrido et al., 2016). The largest residual correlation for a model was a second criteria used to control the amount of population error. The largest population residual correlation for models with a CFI = 0.95 was ≤ 0.10, while for models with a CFI = 0.90 it was ≤ 0.15 (Shi, Maydeu-Olivares, & DiStefano, 2018). This ensured that the amount of population error introduced was within the intended levels for all models.

Importantly, Cudeck and Browne (1992) warned that their method works only when the generated error is not too large. Thereby, to confirm that we achieved a close-fitting population structure for each condition, we fitted a CFA with maximum likelihood using the correct specification of the model, and the resulting CFI was compared with the intended CFI at a tolerance value of 1e-05. Similarly, we also checked whether the estimated parameters were equal to the population parameters and that the largest absolute residuals for each structure was within the intend boundary. In the end, the `sim_factor` function was iterated until a positive definite correlation matrix with error was obtained and satisfied the aforementioned requirements. Finally, for each conditions we extracted 500 samples from a multivariate normal distribution with unit variances and zero means.

Data Analysis

To evaluate the performance of the fit indices, two strategies were employed. The first involves computing balanced accuracy, and the second one involves computing percent correct scores. In simulations one and two, which differ in the number of conditions, the task involves determining whether a correlated traits model or a bifactor model better fits the data. In simulation one, the true data generation mechanism is a correlated traits model, while in simulation two, it is a bifactor model. The *balanced accuracy* was calculated using the *caret* package (version 6.0.94). Balanced accuracy is the average of sensitivity and specificity. The confusion matrix below illustrates these concepts.

	Reference: Correlated Traits	**Reference: Bifactor Model**
Predicted: Correlated Traits	A	B
Predicted: Bifactor Model	C	D

Sensitivity: The proportion of data generated from a correlated traits model that is correctly identified as such, compared to being incorrectly identified as fitting a bifactor model.

$$\text{Sensitivity} = \frac{A}{A + C}$$

Specificity: The proportion of data generated from a bifactor model that is correctly identified as such, compared to being incorrectly identified as fitting a correlated traits model.

$$\text{Specificity} = \frac{D}{B + D}$$

Using balanced accuracy instead of just accuracy has several advantages, particularly in contexts where the number of conditions of the simulations differ, and the accuracy of the fit indices are imbalanced or when it is crucial to give equal importance to both sensitivity and specificity. By taking the average of these two metrics, balanced accuracy ensures that both data generating mechanisms are given equal importance, thus providing a more informative and fair assessment of the fit indices' performance.

In the context of evaluating fit indices for correlated traits models and bifactor models, using balanced accuracy helps ensure that both sensitivity and specificity are considered equally. This leads to a more comprehensive and fair evaluation of the model's performance, especially when the two classes (correlated traits vs. bifactor) may not be equally represented or when the consequences of different types of misclassification are significant. Balanced accuracy has boundaries of zero and one, with zero indicating a very low specificity and sensitivity, and one indicating perfect specificity and sensitivity.

The second strategy uses the percentage of correct structure/model selection (**PC**):

$$PC = \frac{\sum_{i=1}^{N} C}{N}, \text{ for } C = \left\{ \begin{array}{l} 1 \text{ if } S_{selected} = S_{correct} \\ \\ 0 \text{ if } S_{selected} \neq S_{correct} \end{array} \right\}, \tag{11}$$

where N is the number of sample data matrices simulated, $S_{selected}$ is the selected structure/model and $S_{correct}$ is the correct structure/model. The PC criterion has boundaries of 0% and 100%, with 0% signaling complete inaccuracy and 100% indicating perfect accuracy. In the first simulation (correlated traits vs. bifactor structures), the correct structure represents the four correlated traits used in the data-generation process, while the incorrect structure is a bifactor model with four group factors and one general factor. In the second simulation, (bifactor structures vs. correlated traits) the correct structure represents a bifactor model with four group factors and one general factor, while the incorrect is a four-factor correlated traits structure. The fit of the correct structure in the third simulation (see Figure 5) was compared to the fit of six incorrect structures that either had more factors (i.e., overfactored) or less factors (i.e., underfactored) than the correct structure in one or two levels (the number of first-order factors is four per second-order factor; and the number of second-order factors is two): 1) overfactored first-order (one more factor than the correct structure), 2) overfactored second-order (three factors), 3) underfactored first-order (one less factor than the correct structure), 4) underfactored second-order (one factor), 5) underfactored both levels (first and second-order), and 6) overfactored both levels (first and second-order).

The strategy used to select the *best fitting* structure is to use the indices as relative measures of fit. Considering all the conditions, the structure with the highest (CFI) or lowest (RMSEA, SRMR, BIC, AIC, or GenTEFI) fit values is selected. If the selected structure is not the correct one, the respective percent correct score is zero. Percent correct scores of one are only achieved when the selected structure is the correct structure.

We used *R* (version 4.4.4; R Core Team, 2017) for all our analyses. The generalized total entropy fit index was computed using the *EGAnet* package (version 2.0.6; Golino & Christensen, 2022), while CFI, RMSEA, SRMR, BIC, and AIC were calculated using *lavaan* (Rosseel, 2012b). Confirmatory factor models were estimated using maximum likelihood. The figures were generated using the *ggplot2* package (version 3.5.1; Wickham, 2016) and the *ggpubr* package (version 0.6.0; Kassambara, 2018). To compute the *GenTEFI* index, a Pearson correlation matrix was computed and the fit index was calculated using *EGAnet*.

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: https://osf.io/w5ctr/?view_only=3988eafbfcac45e2846fe688f7b7af58.

Results

Simulations One (Correlated Traits vs. Bifactor Structures) and Two (Bifactor Structures vs Correlated Traits): Balanced Accuracy

Figure 6 shows the balanced accuracy at the sample and population level, for lower and higher population error, for all fit indices. Only GenTEFI had very high balanced accuracy both at the population and at the sample level, with scores ranging from .94 (sample level) to .95 (population level), indicating a very high sensitivity and specificity in identifying the correct model as the best-fitting one. GenTEFI was also not affected by the magnitude of the population error, presenting the same balanced accuracy for lower and higher error.

At the population level, RMSEA and SRMR were the next two best performing fit indices, with balanced accuracy of .85 (lower error) and .83 (higher error) for the former, and .67 (lower error) and .64 (higher error) for the latter. AIC, BIC, and CFI presented a balanced accuracy of .54 and .53 at the population level, for lower and higher population error respectively, operating almost at the chance level (a balanced accuracy of .50 for binary classes).

At the sample level, BIC presented a balanced accuracy of .95 for lower error and .91 for higher error, while AIC had a balanced accuracy of .81 and .72, for lower and higher population error, respectively. RMSEA presented a balanced accuracy of .79 and .81, SRMR .62 and .61, and CFI .61 and .57, for lower and higher population error, respectively.

The general results of simulations one and two indicates that the traditional fit indices are not adequate to differentiate a correlated traits structure from a bifactor structure, at both the sample and the population level, for lower or higher population error. BIC, that presents a balanced accuracy similar to GenTEFI at the sample level, suffers at the population level, operating very close to chance. AIC is also operating near chance for the population level, and is negatively impacted by the population error, especially in the sample level. RMSEA presents a moderately high balanced accuracy across error and levels, although markedly smaller than GenTEFI. SRMR and CFI present low balanced accuracies at both the sample and the population levels.

In sum, GenTEFI is the only fit index that is good at capturing DE, DA, and OD, according to the Cudeck and Henly (1991) and Preacher et al. (2013) framework. BIC, on the other hand, is good at capturing DE only, since it presents a high balanced accuracy at the sample level, but its' performance at the population level is poor.

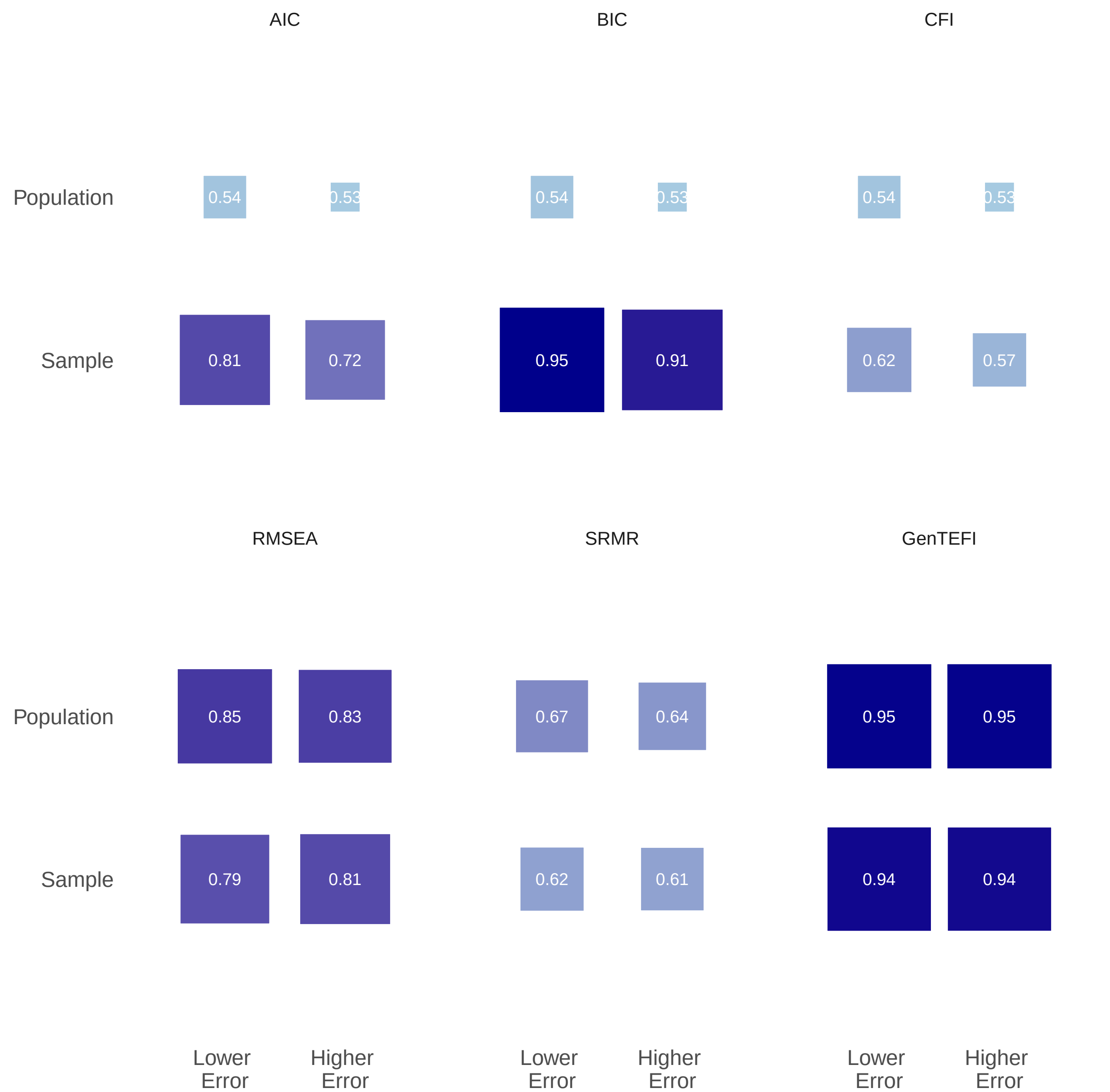


Figure 6

Balanced Accuracy by fit index at the sample and population level, per magniture of population error.

Simulation One (Correlated Traits vs. Bifactor Structures): Percent Correct Scores

In terms of the percent correct scores, GenTEFI (comparing the correlated traits structure with first-order TEFI and the bifactor structure with second-order TEFI from

Equation 4) correctly identified the correlated traits structure as fitting the data better than the bifactor structure in 94.44% of the tested conditions, at the population level, both for the higher and the lower population error conditions. At the sample level, GenTEFI obtained a very similar result, with a mean percentage correct of 94.19% for the higher population error condition, and a mean percentage correct of 94.04% for the lower population error condition.

The second best performing index at the population was the RMSEA, with a mean percent correct of 66.67% for the higher population error condition, and 72.22% for the lower population error condition. However, at the sample level, RMSEA had mean percentage correct of 61.23% for the higher population error condition, and a mean percentage correct of 58.97% for the lower population error condition.

The third best performing index at the population was the SRMR, with a mean percent correct of only 27.78% for the higher population error condition, and 38.89% for the lower population error condition. At the sample level, SRMR performed poorly, with mean percentage correct of 22.45% for the higher population error condition, and a mean percentage correct of 23.06% for the lower population error condition.

AIC, BIC, and CFI performed very poorly at the population level, with the same mean percentage correct for higher and lower population errors of 5.56% and 11.11%. At the sample level, BIC presented the best performance when compared to the other traditional fit indices, with a mean percentage correct of 87.20% for higher population error, and 95.01% for lower population error, followed by AIC (43.96% and 61.74%) and CFI (14.49% and 24.32%).

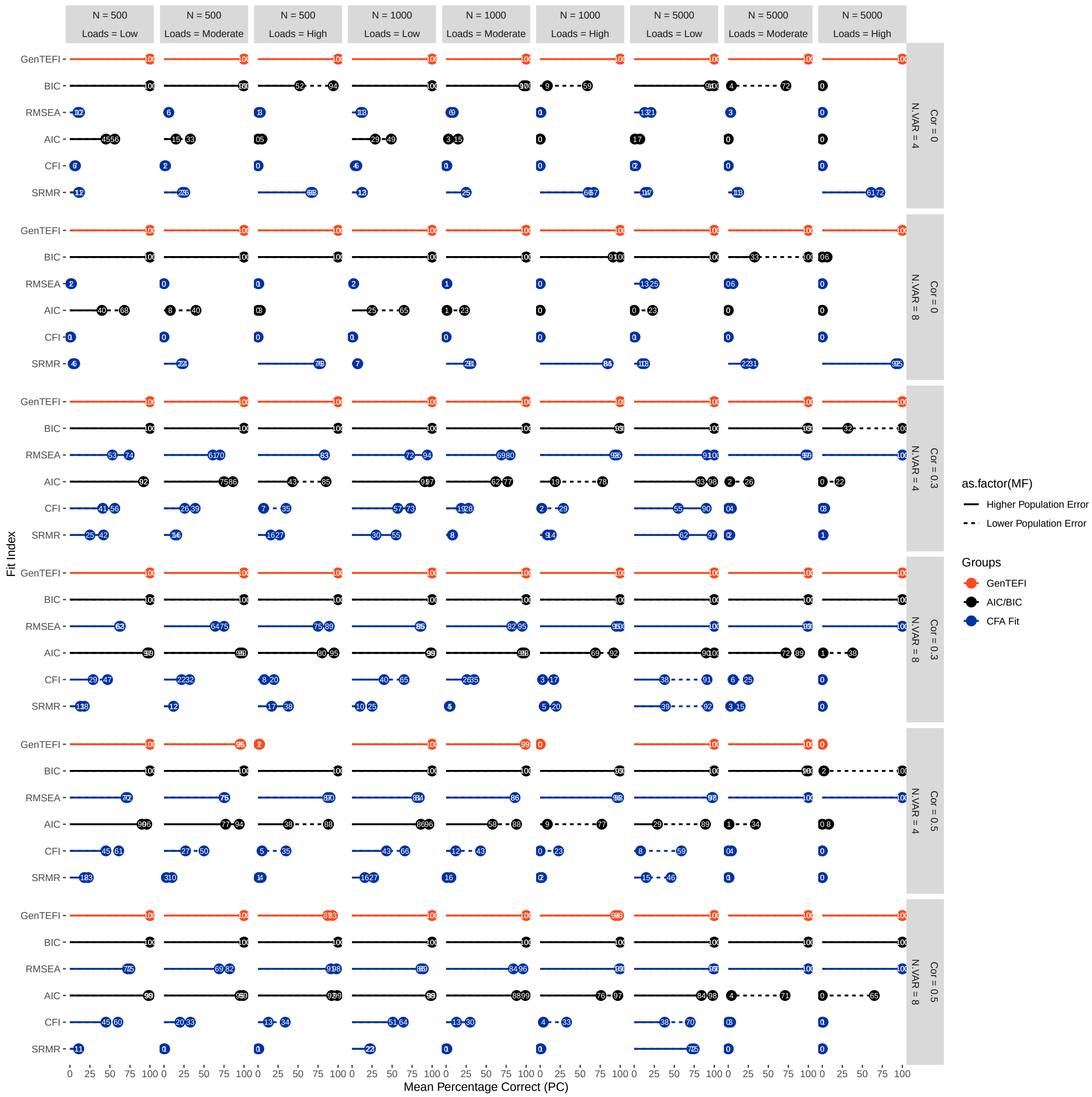


Figure 7

Percentage Correct by fit index comparing a correlated traits structure (true population structure) compared to a bifactor structure. N = sample size, Loads = loadings, Cor = correlations, NVAR = number of variables per factor, and MF = error

Figure 7 shows the mean percentage correct per fit index at different magnitudes of population error, for the sample estimates only, broken by condition (number of variables per factor, number of items per factor, magnitude of loadings, interfactor correlation, and sample size).

Notably, the GenTEFI index has a very high accuracy in almost all conditions tested. The only conditions it presents a very low accuracy occur when the correlations between factors is .5, loadings are high, and there are four items per factor. This result was expected given the property described by Figure 3). The first-order TEFI when compared to the second-order TEFI, clearly shows a lower uncertainty (i.e., lower TEFI values), for conditions with zero and low correlations. For moderate correlations, as factor loadings are high, Figure 3 shows that the uncertainty of a general factor decreases, and that helps to explain why GenTEFI in Figure 7 presented a very low percentage correct the conditions described above. As the items become more strongly related to the factors, with increased correlations between factors, the overall system of variables has a higher entropy or uncertainty than the average entropy of its individual factors. The global interactions of the variables contribute more to its entropy than the internal structure of its factors, and therefore the uncertainty of a hypothetical second-order factor decreases. However, Figure 7 shows that increasing the number of variables per factor addresses the issue caused by the moderate correlations and high loadings, leading GenTEFI to a percentage correct of almost 100%, in most of the conditions tested.

BIC, on the other hand, suffers with high loadings and sample sizes of 1,000 and 5,000, when the correlation between factors is zero, irrespective of the number of items per factor or magnitude of the population error. When the correlation between factors is small or moderate, BIC presents a very high percentage correct, except with sample sizes of 5,000 and higher population error. BIC is negatively impacted by sample size, going from a mean percentage correct of 97.30% for samples of 500 observations to 70.12% for samples of 5,000 observations (in the higher population error conditions compared to 99.66% and 87.63% in the lower population error condition for the same sample sizes).

In sum, the current simulation shows that our new fit index can adequately differentiate a correlated traits structure from a bifactor structure where the traditional fit indices (CFI, RMSEA, SRMR, BIC, and AIC) can't, especially when the sample level and the population level accuracy is taken into consideration. When the data generation mechanism is a correlated traits structure, the GenTEFI index is the only one presenting a high capacity in detecting DE (sample-level) and DA (population level), being the only fit index with an OD detection capacity that is high and not affected by the population error magnitude.

Simulation 2 (Bifactor Structures vs. Correlated Traits): Percent Correct Score

The second simulation compared bifactor structures (reflecting the true data generating mechanism) with correlated traits structures. It is well known that traditional fit indices have a bias towards bifactor models, and therefore it is expected that in the second simulation these fit indices perform well.

At the population level, BIC, AIC, CFI, RMSEA, and SRMR achieved the highest mean percent correct (100%) for both lower and higher population error. GenTEFI had a mean percentage correct of 95.83% for both error levels at the population.

At the sample level, CFI, AIC, and SRMR presented a mean percentage of 100% for higher and lower population error, respectively, while RMSEA presented percent correct scores of 99.98% and 100%, for higher and lower error. BIC presented percent correct scores of 95.23% and 95.94%, for higher and lower population errors, while GenTEFI presented percent correct scores of 93.98% and 94.46%.

In the second simulation, with a true data generation mechanism being a bifactor model, all fit indices presented a high or very high accuracy for the population and sample levels. This indicates that all fit indices are good at capturing DE, DA, and OD when the true data generation mechanism is a bifactor model.

Simulation Three (Bifactor Structures vs Bifactor Structures): Percent Correct

At the population level, when the correct bifactor structure was compared to incorrect bifactor structures (over or underfactoring the group or general factors), RMSEA presented the higher percent correct score, with 97.92% for lower and 94.44% for higher population error. The second most accurate fit index at the population level was SRMR with percent correct scores of 86.81% and 86.11%, for lower and higher population error, respectively. BIC presented percent correct scores of 83.33% and 85.42%, closely tied with AIC (83.33% and 83.33%), and CFI (83.33% and 83.33%), for both lower and higher population error. GenTEFI presented percent correct scores of 80.56% and 79.86%, for lower and higher population error.

At the sample level, the percent correct scores changed for both higher and lower population error, with BIC presenting a PC of 98.18% and 98.04%, AIC a PC of 92.88% and 93.31%, RMSEA a PC of 96.03% and 93.87%, CFI a PC of 89.54% and 89.92%, and SRMR a PC of 86.6% and 86.53%, respectively for higher and lower error. GenTEFI presented a PC of 80.24% and 80.32%, for higher and lower error at the sample level.

When breaking down the population level accuracy by type of structure, the traditional fit indices presented a percent correct score of 100% for all structures, except for the *Over GF* (overfactored general factors). In this condition, AIC and CFI presented a percent score of 0%, while BIC presented a PC of 6.25%, SRMR a PC of 18.75%, and RMSEA a PC of 77.08%. In the *Over GF* structure, GenTEFI presented a PC of 70.83%, 68.75% in the *Over 1st*, 75% in the *Both Over*, 95.83% in the *Under GF*, 78.16% in the *Under 1st*, and 91.66% in the *Both Under* structures.

A similar pattern was obtained at the sample level, with all traditional fit indices presenting a PC of (almost) 100% for all structures except *Over GF*. SRMR presented a PC of 19.58%, while CFI presented a PC of 38.38%, RMSEA of 69.72%, AIC of 58.58%, and

Table 3

Conditions presenting a large effect size for Simulation 3.

	CFI	RMSEA	SMRS	AIC	BIC	GenTEFI
Structure	0.61	0.34	0.78	0.45	0.17	0.21
VAR.GRF:LOAD.GF	0.00	0.00	0.00	0.00	0.00	0.40
VAR.GRF:Structure	0.00	0.00	0.02	0.01	0.00	0.49
LOAD.GF:Structure	0.00	0.01	0.00	0.00	0.01	0.54
COR.GF:Structure	0.00	0.00	0.00	0.00	0.00	0.18

Note. Effect size computed as partial eta squared from the ANOVA results. The dependet variable = Hit Rate.

BIC of 88.69%. In the *Over GF* structure at the sample level, GenTEFI presented a PC of 72.07%, 70.76% in the *Over 1st*, 76.38% in the *Both Over*, 95.41% in the *Under GF*, 79.28% in the *Under 1st*, and 87.85% in the *Both Under* structures.

Figure 8 breaks down the percent correct scores per number of variables per group factor, magnitude of loadings in the general factors, as well as correlations between the general factors for all overfactoring and underfactoring structures. These conditions presented a large effect size for at least one fit index (see Table 3).

For the *Over GF* structure, the traditional fit indices showed a higher variability in the percent correct scores. CFI, AIC, and SRMR presented PC scores ranging from very low to moderate. RMSEA presented PC scores from low (46%) to moderately high (85%), and BIC presented PC scores from moderate (77%) to very high (99%). GenTEFI presented very high PC scores for moderate loadings in the general factor (100%), but low (15%) to moderately high (81%) PC scores when the loadings in the general factor was low.

For the *Over 1st* structure, all fit indices presented a perfect percent correct score when the number of variables per group factor was four. With eight variables per group factor, GenTEFI presented a low or very low accuracy when the loadings in the general factors was low, but very a very high percent correct score when the loading was moderate (except when the correlation between general factors was .30, condition in which GenTEFI performed poorly even with moderate loadings in the general factors). A similar pattern was observed for the *Both Over* structure.

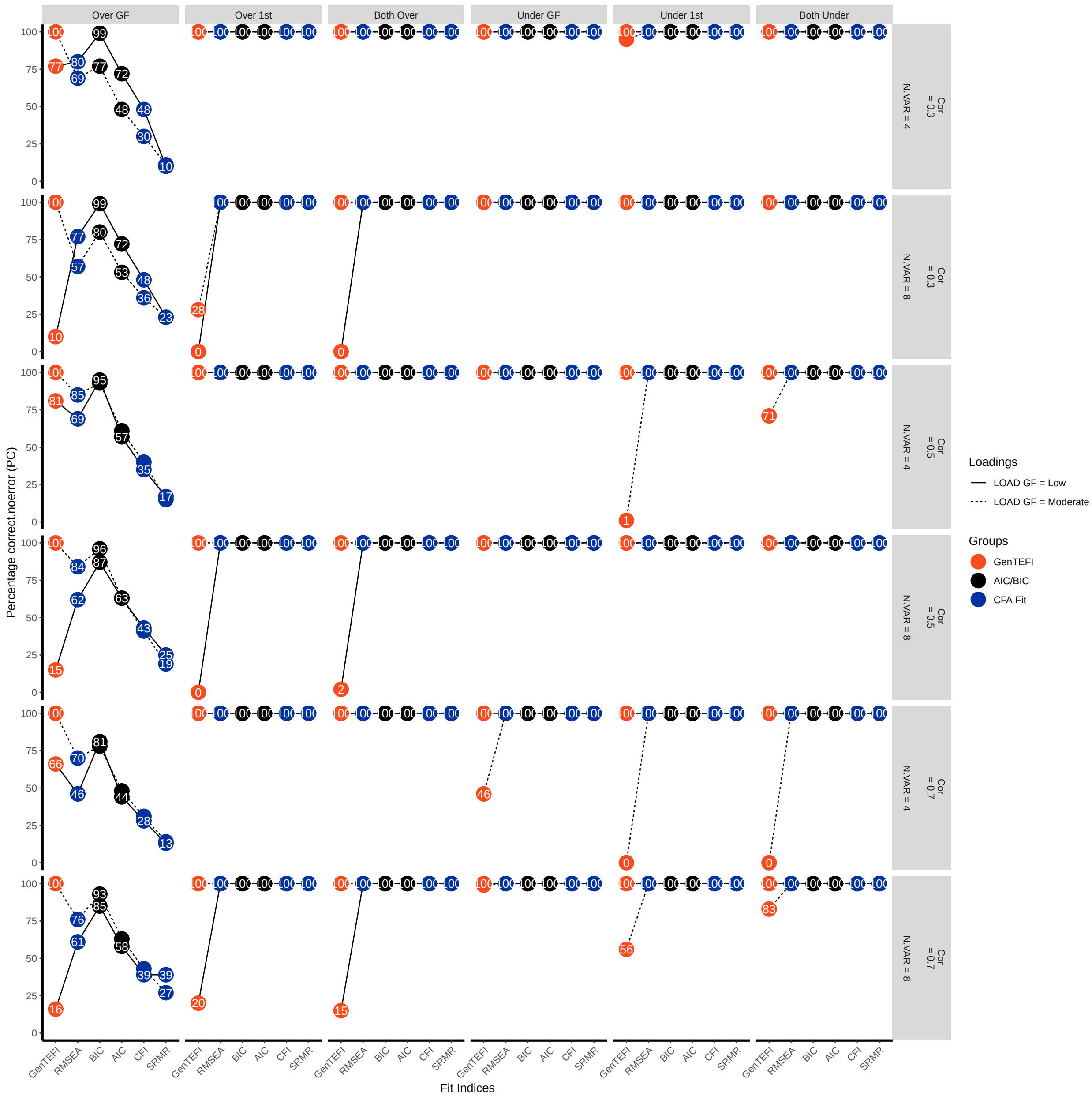


Figure 8

Percentage Correct by fit index comparing a correct bifactor structure with an incorrect bifactor structure (overfactoring the first or the second order factors).

For the underfactored structures, the traditional fit indices presented a perfect percent score for all conditions tested. GenTEFI also presented a perfect PC score in almost all conditions with eight variables per group factor, but presented low to moderate PC scores when the number of variables per group factor was four, the loadings in the general factor was moderate, and the correlation between general factors was .50 and .70.

Empirical Examples

In this section, two empirical examples are briefly described. The first one uses the *Experiences in Close Relationships* scale (Lafontaine et al., 2015), that illustrates well data that fits a structure akin to a correlated traits structure. The second example uses the Misinformation Susceptibility Test (MIST) was published (Maertens et al., 2023), illustrating data that fits better a structure akin to a bifactor structure.

Example One: Experiences in Close Relationships

We analyzed data from the 12-item Experiences in Close Relationships [*ECR-12*; Lafontaine et al. (2015)] scale to compare the fit of a correlated-factors model against a bifactor model. The ECR-12 is purported to measure two attachment factors, anxiety and avoidance, each measured by six items. In general, the ECR published literature has not considered a bifactor model of attachment for the responses to this scale. The data was obtained from the Open Source Psychometrics Project (openpsychometrics.org), which contained full responses by 27,883 persons from the United States.

The items were responded via a 5-point Likert scale with options strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), and strongly agree (5). Cronbach's alpha reliability was .858 for avoidance and .840 for anxiety.

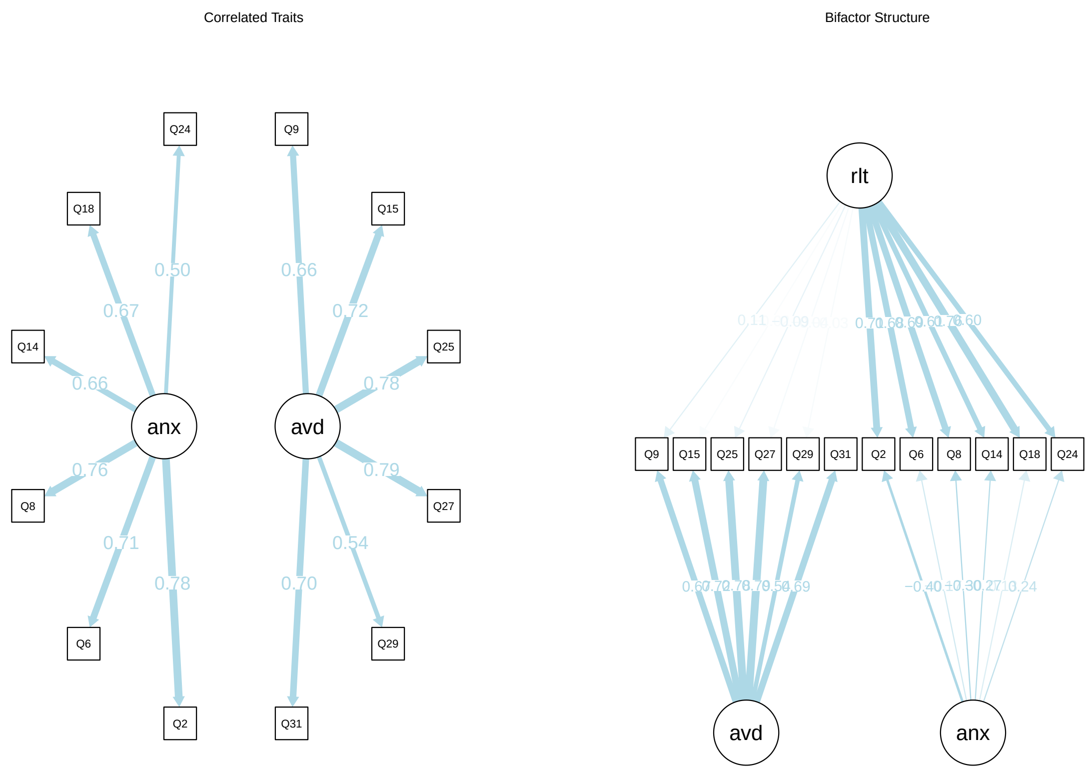


Figure 9

Correlated traits (left) and Bifactor (right) structures of the Experiences in Close Relationships Scale.

The factor models were estimated using robust maximum likelihood. The two-factor CFA model produced a good fit to the data with CFI = 0.95, RMSEA = 0.07, and SRMR = 0.05. Additionally, AIC = 1,825,558.68 and BIC = 1,825,779.91 for this model (see left plot on Figure 9). The standardized loadings on the avoidance factor ranged from .54 to .80 (M = .70), while the loadings on the anxiety factor ranged from .50 to .79 (M = .6785). The two attachment factors were estimated to be approximately orthogonal, with a correlation of -0.002. The bifactor model produced an even better fit to the data according to all factor-analytic fit indices, with CFI = 0.97, RMSEA = 0.05, and SRMR = 0.03, AIC = 1,819,293.09, and BIC = 1,819,611.66 for this model (see right plot on Figure 9).

However, the standardized loadings did not support this bifactor structure as none item of the avoidance dimension had loadings above .12 on the general factor (M = -0.01), while the anxiety items had a mean loading of .67 on it. Conversely, the anxiety items did not load on the anxiety group factor well with a mean loading of -0.13, while the avoidance items had a large mean loading of .70 on the avoidance group factor.

When Exploratory Graph Analysis in the ECR data, the two-dimensional structure of the avoidance and anxiety items became very clear (see Figure 10). The generalized total entropy fit results point to the first-order two-community structure with avoidance and anxiety factors ($TEFI_{first-order}$ = -9.74) as fitting better than the bifactor structure with a general factor ($TEFI_{second-order}$ = -6.22). The lower TEFI value for the first-order structure, compared to the second-order TEFI value, indicates that a two-community structure presents a lower level of disorganization, fitting the data better.

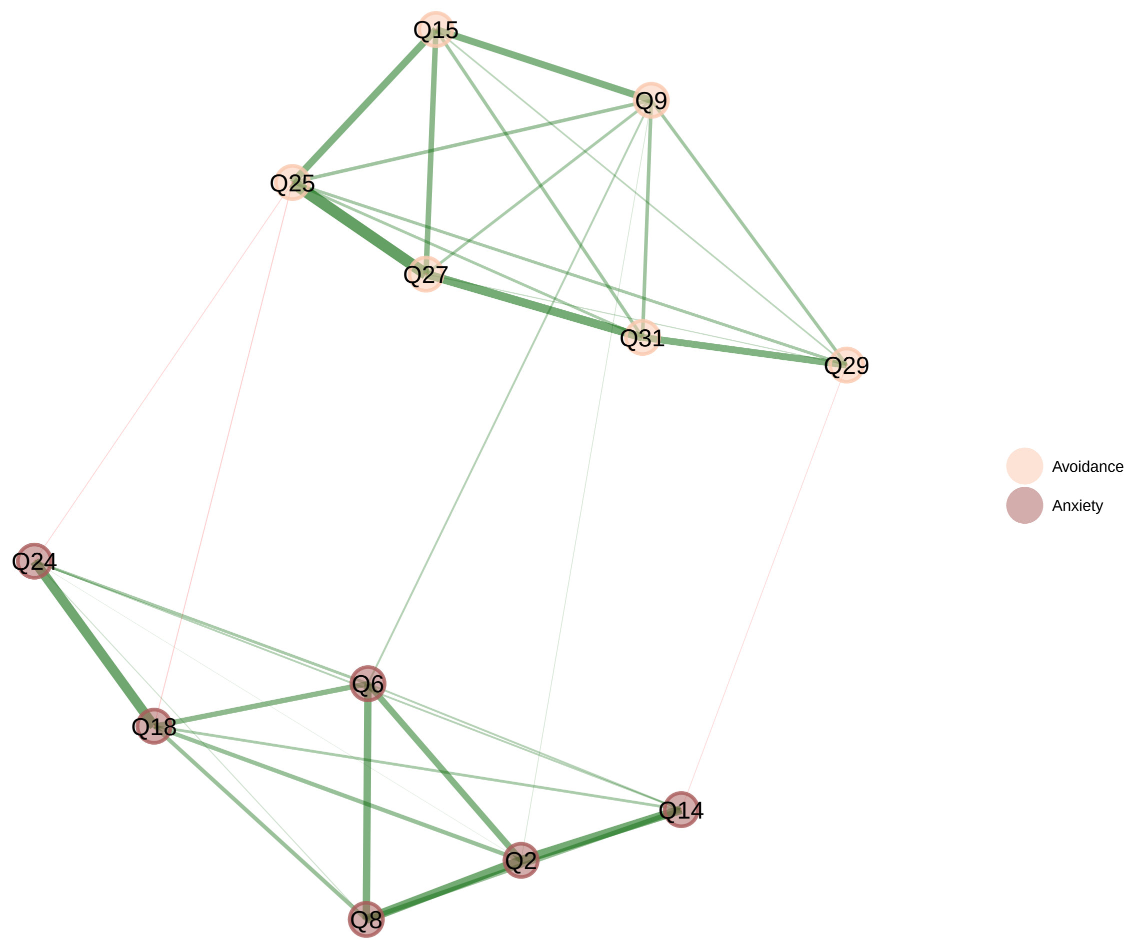


Figure 10

Exploratory Graph Analysis with Walktrap community detection algorithm

The two-factor structure aligns closely with attachment theory as developed by Bowlby and Ainsworth. This theory conceptualizes attachment in terms of anxiety about abandonment and avoidance of intimacy. Lafontaine et al. (2015) points that the two-dimensional factor structure of the ECR has been observed consistently across various populations, cultures, and languages. This robust replication supports the validity of the two-factor model. The paper emphasizes that anxiety and avoidance are designed to be orthogonal (uncorrelated) dimensions. This independence is theoretically important and has been supported by research, although small correlations are sometimes observed. Therefore, a bifactor model does not make theoretical sense, altought it is shown to fit better the data than a correlated-traits structure with two factors, according to the traditional indices (CFI, RMSEA, SRMR, AIC, and BIC). The new Generalized Total Entropy Fit index, on the other hand, identified the two-factor structure as fitting the data better than a the bifactor structure.

Example Two: The Misinformation Susceptibility Test

Recently, a series of three studies aimed at developing and validating a comprehensive measurement tool for misinformation susceptibility called the Misinformation Susceptibility Test (MIST) was published (Maertens et al., 2023). The researchers created a set of fake news headlines (using *GPT-3*) and real news headlines from trusted sources. They employed several psychometric analyses, including EGA (Golino & Epskamp, 2017; Golino et al., 2020) and hierarchical EGA (Jiménez, Abad, Garcia-Garzon, Golino, et al., 2023), to select the items and investigate the structural validity of the instrument.

The studies collectively demonstrate the feasibility of developing a psychometrically validated measurement instrument for misinformation susceptibility. The MIST showed evidence of superiority in predicting outcomes compared to other measures. Overall, the Misinformation Susceptibility Test (MIST) provides a reliable and standardized tool for assessing individuals' susceptibility to misinformation, allowing researchers to gain insights into different facets of misinformation discernment and judgment biases.

In the current paper, Sample 2E from the MIST paper (Maertens et al., 2023) is used. The reason to use this sample is to contrast and replicate the hierarchical EGA of the original paper (used in Sample 2E) with the analysis implemented in this section. Sample 2E was collected in November 2022 using Respondi/Bilendi, using the 16 items version of the MIST instrument. The sample is a nationally representative quota sample (N = 1213) of adults from the US, with 54% of the participants identifying as female (44% male, 2% nonbinary), 33% aged between 18 and 34 years, 31% between 35 and 54 years, and 36% between 55 and 75 years; 24% of the participants reported coming from the Midwest, 17%

from the Northeast, 40% from the South, and 20% from the West of the US.

Hierarchical EGA (*hierEGA*) was implemented in the current study using the *EGAnet* package (version 2.0.6; Golino & Christensen, 2022) using two different strategies. First, *hierEGA* is implemented using the *Walktrap* algorithm to identify the group factors (first-order factors or first-order communities). In the second strategy, the *Louvain* algorithm is used to identify the group factors (or first-order factors). The resulting bifactor structure for each *hierEGA* estimation approach is compared using the generalized total entropy fit index.

```
### Load Data:

data.disinfo <- read.csv2("MIST_Data_Golino_2022.csv")



#### MIST Scale:



mist.items <- colnames(data.disinfo)[7:21]

# False Items
mist.false <- as.data.frame(apply(data.disinfo[,
c(mist.items[c(1:3,6,10,12,14)])],
2,function(x)ifelse(x == 1,1,0)))

# True Items
mist.true <- as.data.frame(apply(data.disinfo[,
c(mist.items[c(4,5,7,8,9,11,13,15)])],
```

```
2,function(x)ifelse(x == 2,1,0)))


# Merge the items to the dataset
disinfo.all <- cbind.data.frame(data.disinfo[,1:6],
                                mist.false, mist.true)




## Hierarchical EGA ANALYSIS:
library(EGAnet)


# Louvain Algorithm (Community Detection)
set.seed(1234)
hierega.disinfo <- hierEGA(disinfo.all[,-c(1:6)],
                           scores = "network",
                           consensus.iter = 1000,
                           consensus.method = "most_common",
                           lower.algorithm = "louvain"
)


# Walktrap Algorithm (Community Detection)
set.seed(1234)
hierega.disinfo.walktrap <- hierEGA(disinfo.all[,-c(1:6)],
scores = "network",
lower.algorithm = "walktrap"
)
```

```
# Compute GenTEFI:
gentefi.hierega <- genTEFI(hierega.disinfo)$VN.Entropy.Fit



gentefi.hierega.walk <- genTEFI(hierega.disinfo.walktrap)$VN.Entropy.Fit

# Alternatively, use the correlation matrix and a list
# representing the lower and the higher order structure:


gentefi.hierega <- tefi(hierega.disinfo$lower_order$correlation,
structure = list(lower_order =
hierega.disinfo$lower_order$wc,
higher_order = hierega.disinfo$higher_order$wc))

gentefi.hierega.walk <- tefi(hierega.disinfo.walktrap$lower_order$correlation,
structure = list(lower_order =
hierega.disinfo.walktrap$lower_order$wc,
higher_order =
hierega.disinfo.walktrap$higher_order$wc))
```

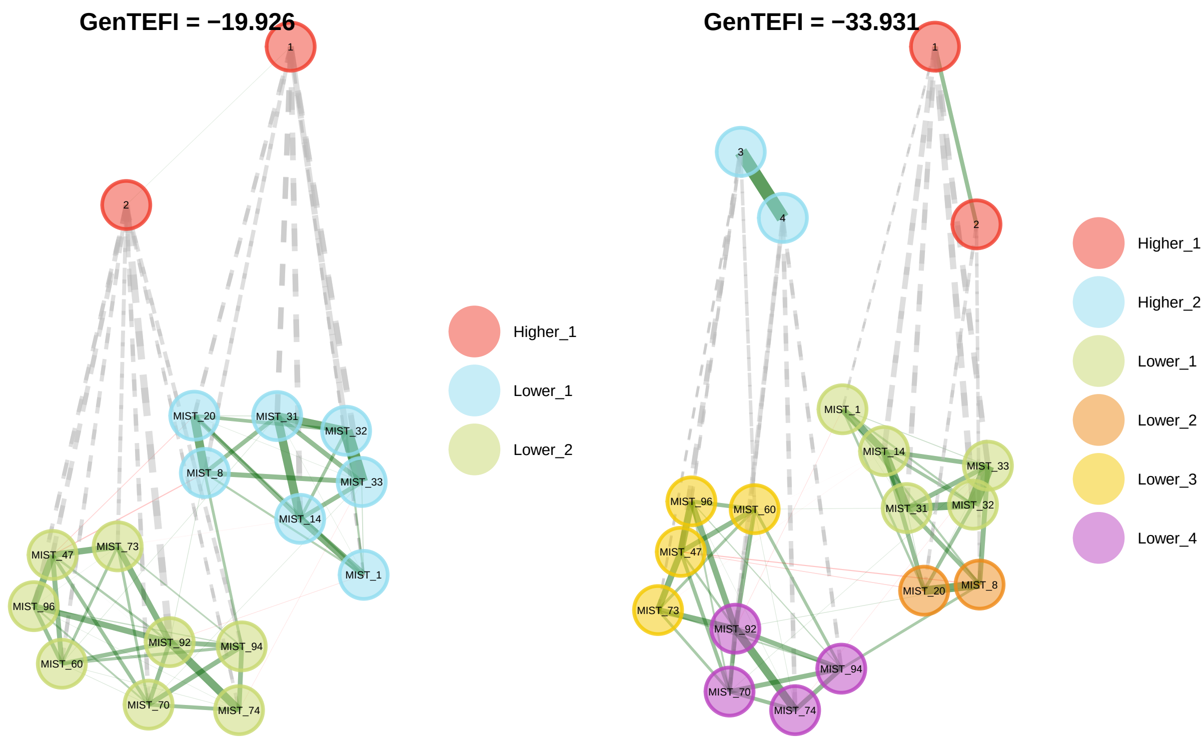


Figure 11

Hierarchical exploratory graph analysis with the Walktrap (left) and Louvain (right) community detection algorithms and the resulting generalized total entropy fit index value for each structure.

Figure 11 shows that the hierarchical EGA technique with the Walktrap algorithm identified only two first-order dimensions (real and fake news items), and one general second-order dimension. In contrast, the hierarchical EGA technique with the Louvain algorithm identified the original four-dimensional (first-order) structure and two general factors found by Maertens et al. (2023) when analyzing the same data. The four first-order structure with two general dimensions (right side of Figure 11) estimated using hierarchical EGA with the Louvain algorithm resulted in the lowest GenTEFI value, fitting the data better than the two group-factor structure with one general factor estimated using the Walktrap algorithm. Additionally, the second-order TEFI (-20.32) with the first-order TEFI (-13.61) of the four first-order structure with two general dimensions (right side of Figure 11), the former was less than the latter, indicating that the two-level organization presents lower uncertainty than the four-factor correlated traits structure (first-order structure only).

Dimension one (green nodes on Figure 11) has fake news items related to general conspiracy beliefs, such as item MIST 1 (A Small Group of People Control the World Economy by Manipulating the Price of Gold and Oil), and conspiracies related to the government, such as items MIST 31 (The Government Is Actively Destroying Evidence Related to the JFK Assassination) and MIST 32 (The Government Is Conducting a Massive Cover-Up of Their Involvement in 9/11). Dimension two (orange nodes on Figure 11) has fake news items about science, such as item MIST 8 (Climate Scientists' Work Is "Unreliable", a "Deceptive Method of Communication"), and false statements against people with a liberal world view, such as items MIST 16 (Left-Wingers Are More Likely to Lie to Get a Good Grade) and MIST 20 (New Study: Left-Wingers Are More Likely to Lie to Get a Higher Salary). Dimensions one and two are part of a second-order dimension of *fake news*.

Dimension three (yellow nodes on Figure 11) is a combination of US and international real news headlines, with items such as MIST 92 (Taiwan Seeks to Join Fight Against Global Warming), and MIST 60 (Hyatt Will Remove Small Bottles from Hotel Bathrooms by 2021). Dimension four (purple nodes on Figure 1) has real news items related to politically charged topics in the US, such as items MIST 70 (Majority in US Still Want Abortion Legal, with Limits), MIST 74 (Most Americans Say It's OK for Professional Athletes to Speak out Publicly about Politics), and MIST 94 (United Nations Gets Mostly Positive Marks from People Around the World). Dimensions three and four are part of a second-order dimension of *real news*.

Discussion

In this study, we introduced the Generalized Total Entropy Fit Index (GenTEFI), a novel tool to assess the fit of competing dimensionality structures in psychometric research. GenTEFI extends the recently developed Total Entropy Fit Index (Golino et al., 2021) to accommodate more complex, two-level structures such as those represented by bifactor models with multiple correlated general factors. Our simulation studies demonstrate that GenTEFI is the only fit index that can adequately differentiate between correlated traits and bifactor structures in both the sample and population level (simulations one and two), addressing a critical limitation in the field.

Fit indices are crucial in psychometrics as they allow researchers to evaluate how well a statistical model fits the empirical data, thereby aiding the interpretation of the underlying structure of the variables and making predictions about their behavior (Golino et al., 2021; Savalei & Rhemtulla, 2013; Shi, DiStefano, McDaniel, & Jiang, 2018). However, traditional fit indices have limitations, such as assumptions about the data and model, and the use of cut-off values with low power to detect population structure in simulation studies. Additionally, recent simulation studies have pointed to the strong bias of traditional fit indices when comparing correlated traits and bifactor structures, usually suggesting the latter as fitting data better (Bonifay & Cai, 2017; Greene et al., 2019; Morgan et al., 2015; Murray & Johnson, 2013).

Examining the limitations of the traditional fit indices in this context, the field has adopted two general approaches that we call *evasion* and *bifactor avoidance*. The evasive approach argues that other elements should be used when comparing different structures in psychological research, such as theory and reliability, and that reliance on fit indices is inadequate. The bifactor avoidance approach, on the other hand, argues that bifactor structures should not be used because they overfit the data. In this paper, we argue that both positions overlook the main issue: A new fit index that can be used to compare the fit of competing dimensionality structures, such as a correlated traits model and a bifactor model, is necessary. In response to the limitations of the traditional fit indices, our study introduced the *GenTEFI*.

Our study also advances previous simulation studies investigating the accuracy of traditional fit indices to differentiate between bifactor and other competing models. These previous simulations focused solely on DE, which quantifies sampling error or the discrepancy between population and sample estimates (Cudeck & Henly, 1991; Preacher et al., 2013). In our simulations, we use (Cudeck & Henly, 1991) different types of discrepancy. DA represents the difference between the population covariance matrix and the model's implied covariance matrix in the population. DA is a measure of the model's lack of fit in the population, independent of sample size and sampling variability, and was implemented in our study by computing the accuracy of different fit indices at the population level, with two magnitudes of errors: lower and higher population error. It quantifies the degree of model misspecification or the extent to which the model deviates from the true data-generating process. Minimizing DA is equivalent to maximizing the model's verisimilitude or proximity to the truth.

Previous research showed that traditional fit indices are biased in favor of bifactor models (Bonifay & Cai, 2017; Greene et al., 2019; Morgan et al., 2015). Similarly, Kan et al. (2024) argued that approximate fit indices (like CFI and RMSEA), which reflect sample discrepancy between the sample covariance matrix and the model-implied covariance matrix, should not be used for comparing models. Instead, they proposed that relative fit metrics like AIC and BIC, which approximate the OD, are more accurate in differentiating bifactor models from other models. However, as stated above, these studies share a common limitation: they focus solely on the accuracy of fit indices in selecting the model with the highest generalizability (i.e., minimizing OD), failing to consider the equally important goal of selecting the model with the highest verisimilitude (i.e., minimizing DA).

When the emphasis is placed on a single type of evidence, due to the data generation mechanism only capturing model divergence resulting from estimation (i.e., the data is generated from a model-implied covariance matrix that perfectly represents the population covariance matrix), it becomes impossible to evaluate the fit indices' ability to detect divergence caused by approximation. Divergence due to approximation can only occur when the data generation mechanism's model-implied covariance matrix does not perfectly match the population covariance matrix. In other words, introducing error into the population is necessary to capture more realistic evidence of divergence caused by estimation (at the sample level) and to capture divergence due to approximation (at the population level). Previous studies did not include errors in the population and did not explore how various fit indices perform at the population level (Bonifay & Cai, 2017; Greene et al., 2019; Kan et al., 2024; Morgan et al., 2015).

The result of the present study shows that *GenTEFI* is the only fit index that can adequately differentiate a correlated traits structure from a bifactor structure in terms of fit both at the population (DA) and sample levels (DE), irrespective of the true data generation mechanism. Considering all conditions tested in the first and second simulation of our study, *GenTEFI* presented a balanced accuracy of 94% at the sample level and 95% at the population level for both lower and higher population error. *GenTEFI* is a very good fit index to capture both discrepancy due to estimation and discrepancy due to approximation, and it was not impacted by the magnitude of population error. BIC and AIC, pointed out by Kan et al. (2024) as the go-to fit indices for comparing bifactor models with competing models, presented a very high and moderately high balanced accuracy, respectively, at the sample level. However, both fit indices performed poorly at the population level, indicating that their capacity to detect DA is inadequate. In the population, BIC and AIC favor the larger model if there is population error, and they are impacted by the magnitude of the population error, although the latter is more than the former. RMSEA and SRMR presented a moderately high and moderate balanced accuracy at the sample and the population level but with performances much lower than the performance observed by *GenTEFI*.

When considering the results of simulation one (correlated traits vs. bifactor) separately, we observed that *GenTEFI* was the most effective fit index to differentiate a correlated traits from a bifactor structure when the data generation mechanism was, in fact, a correlated traits model. GenTEFI presented low percentage correct values when the correlation between factors was .5, the loadings were high, and the number of items per factor was four. With high loadings and increased factor correlation, the uncertainty of a general factor decreases because, in this condition, the overall system of variables has a higher entropy or uncertainty than the average entropy of its individual factors. The global interactions of the variables contribute more to its entropy than the internal structure of its factors, and therefore, the uncertainty of a hypothetical second-order general factor decreases. However, increasing the number of variables per factor addresses the issue caused by the moderate correlations and high loadings, leading GenTEFI to a percentage correct above 87%. Our results show that BIC is more conservative the smaller the sample size, thus will favor the smaller model (correlated traits model) with smaller samples. With increased sample sizes, BIC performs worse, favoring the larger (bifactor) model. When the data came from a bifactor model, then the fit indices performed as expected, and all presented percent correct scores above 90%. This result aligns with the most recent evidence (Bonifay & Cai, 2017; Greene et al., 2019; Morgan et al., 2015, 2015) showing that traditional fit indices tend to be biased toward bifactor structures over correlated factor structures.

In the third simulation, the results indicate that RMSEA was the most accurate fit index in identifying the correct bifactor structure when comparing it to several misspecified structures that either underestimated or overestimated the number of group and/or general factors, both at the population and at the sample level. All fit indices presented percent correct scores around or above 80% at the population level and above 90% at the sample level. This indicates that the fit indices are very good at identifying the correct bifactor structure compared to misspecified group or general factor structures. The “Over GF” structure was the most difficult structure misspecification tested in the third simulation since it overestimates the number of general factors. Overall, *GenTEFI*, *BIC*, and *RMSEA* presented the highest percent correct scores in this condition and worked well or very well when the loadings in the general factor were moderate.

It is interesting to note that the traditional fit indices work well in the “Over First” condition but not in the “Over GF.” When the number of first-order factors increases, compared to the number of group factors in the true population model, it does not add any factor correlations to the model because these are orthogonal group factors in the bifactor model. Whereas when one additional general factor is added in the "Over GF" condition, several parameters are added in the new factor correlations. Thus, "Over GF" actually implies a larger model than the true bifactor model used to generate the data, and the traditional fit indices have a tendency to favor larger, more complex models. *GenTEFI*, on the other side, identifies the correct structure in the "Over GF" condition perfectly when the loadings in the general factors are moderate. Items that are more informative with respect to the general factors are all that *GenTEFI* needs, as shown by the results of our third simulation. The issue of the number of parameters also helps explain why traditional fit indices tend to select a bifactor model even when the data is generated from a correlated traits model. The bifactor model posits additional loading parameters for each item on the general factor, thus constituting a notably larger model than the correlated traits one.

Taking the results across simulations together, there are a clear set of recommendations that applied researchers should follow. First, based on simulations 1 and 2, *GenTEFI* should be the sole metric used to decide between a correlated traits or bifactor model. No other fit index achieved results that would satisfy a balanced, unbiased estimate of the generating model. If *GenTEFI* suggests a bifactor model, then, based on simulation 3, traditional fit indices (e.g., *RMSEA*, *BIC*) and *GenTEFI* are acceptable and can be used in combination to determine the appropriate structure among competing bifactor structures. Undoubtedly, the first step is the most crucial one.

Conclusion

The development of GenTEFI represents a significant advancement in psychometric research. By leveraging information theory and quantum information theory, GenTEFI provides a principled and objective criterion to evaluate dimensionality structures in two different levels (first and second-order). Unlike traditional fit indices, which rely on the distance or equivalence between model-implied and observed covariance matrices, GenTEFI captures the reduction in uncertainty or disorder achieved by partitioning a multidimensional space into subsets of variables in two different levels. This entropy-based approach offers a fresh perspective on model fit that is less sensitive to the biases and limitations of conventional methods.

In practice, GenTEFI can be a valuable tool for researchers investigating the structure of psychological constructs. It is particularly advantageous when comparing correlated traits and bifactor models, a common challenge in fields such as personality, intelligence, and psychopathology. By accurately identifying the underlying structure, GenTEFI can help researchers develop more valid and reliable measurement instruments and advance theoretical understanding. Moreover, GenTEFI's ability to handle complex, hierarchical structures makes it suitable for a wide range of research questions and data types.

Our simulation studies provide compelling evidence for the superiority of GenTEFI in certain contexts. When the true structure follows a correlated traits model, GenTEFI consistently outperforms traditional fit indices in selecting the correct structure; when the true structure follows a bifactor structure, GenTEFI exhibits high accuracy in detecting the best fitting structure. GenTEFI also presented a moderately high accuracy in the third simulation comparing bifactor with bifactor structures with different types of misspecifications, such as underfactoring or overfactoring at different levels. However, it is important to acknowledge that there may be conditions under which researchers might prefer conventional fit indices. For instance, if the primary goal is to assess the absolute fit of a single model rather than compare competing structures, indices like RMSEA or CFI may be more appropriate.

Our empirical examples provide valuable insights into model selection. In a large dataset using the *Experiences in Close Relationships Questionnaire*, traditional fit indices (CFI, RMSEA, SRMR, AIC, and BIC) favored a bifactor model over a two-factor model of attachment styles in close relationships. However, the Generalized Total Entropy Fit Index indicated a superior fit for the two-factor structure. This aligns with the strong theoretical and empirical support for the two-factor structure, as well as its practical utility and interpretability in conceptualizing adult attachment (Lafontaine et al., 2015). It's noteworthy that Lafontaine et al. (2015) did not directly compare the two-factor and bifactor models. Such comparisons often encounter the well-documented issue of fit indices being biased towards bifactor models, which contradicts both attachment theory and the patterns observed in ECR data when using Exploratory Graph Analysis (as shown in our paper).

Conversely, our second empirical example demonstrates that the *Misinformation Susceptibility Test* is better represented by first and second-order communities/factors, as evidenced by the Generalized Total Entropy Fit Index results. This finding aligns well with the theoretical framework of misinformation susceptibility. These contrasting examples highlight the importance of considering both statistical indicators and theoretical foundations when selecting appropriate models for psychological constructs.

Future research could explore several promising directions to further refine and extend GenTEFI. One avenue is to investigate its performance under a broader range of data conditions, such as different sample sizes, variable distributions, and model complexities. Additionally, integrating GenTEFI with other advanced psychometric techniques, such as exploratory structural equation modeling, could yield novel insights into the structure of psychological phenomena.

By addressing the limitations of traditional fit indices and providing a theoretically grounded approach based on entropy reduction, GenTEFI offers researchers a powerful new tool for understanding the structure of psychological constructs. Our findings underscore the potential of GenTEFI to enhance the validity and reliability of measurement instruments and contribute to theoretical advancements in psychology. As researchers continue to grapple with the challenges of modeling complex, hierarchical structures, GenTEFI provides a promising avenue for future research and application.

References

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

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

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

Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255–265.

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

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

Bonifay, W., & Cai, L. (2017). On the complexity of item response theory models. *Multivariate Behavioral Research*, *52*(4), 465–484.

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, *5*(1), 184–186.

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

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, *75*(4), 581–612. https://doi.org/10.1007/s11336-010-9178-0

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

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

Christensen, A. P., & Golino, H. (2021). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A monte carlo simulation and tutorial. *Psych*, *3*(3), 479–500.

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

Cucina, J., & Byle, K. (2017a). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence*, *5*(3), 27.

Cucina, J., & Byle, K. (2017b). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence*, *5*(3), 27. https://doi.org/10.3390/jintelligence5030027

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

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the" problem" of sample size: A clarification. *Psychological Bulletin*, *109*(3), 512.

https://doi.org/10.1037/0033-2909.109.3.512

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*, 541–562. https://doi.org/10.1037/met0000083

Eid, M., Krumm, S., Koch, T., & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal of Intelligence*, *6*(3), 42. https://doi.org/10.3390/jintelligence6030042

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

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

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

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57–68. https://doi.org/10.1016/j.intell.2016.01.006

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., & Christensen, A. P. (2022). *EGAnet: Exploratory Graph Analysis – A framework for estimating the number of dimensions in multivariate data using network psychometrics.*

Golino, H., Moulder, R., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., … Boker, S. M. (2021). Entropy fit indices: New fit measures for assessing the structure and dimensionality of multiple latent variables. *Multivariate Behavioral Research*, *56*(6), 874–902.

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

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., et al.others. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology*, *128*(7), 740.

Hall, B. C. (2013). *Quantum theory for mathematicians* (Vol. 267). Springer.

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

Hu, L.-T., & Bentler, P. M. (1999). 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

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

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory Bi-factor Analysis: The Oblique Case. *Psychometrika*, *77*(3), 442–454. https://doi.org/10.1007/s11336-012-9269-1

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

Jiménez, M., Abad, F. J., Garcia-Garzon, E., & Garrido, L. E. (2023b). Exploratory bi-factor analysis with multiple general factors. *Multivariate Behavioral Research*, *58*(6), 1072–1089.

Jiménez, M., Abad, F. J., Garcia-Garzon, E., & Garrido, L. E. (2023a). Exploratory bi-factor analysis with multiple general factors. *Multivariate Behavioral Research*, *58*(6), 1072–1089.

Jiménez, M., Abad, F. J., Garcia-Garzon, E., Golino, H., Christensen, A. P., & Garrido, L. E. (2023). Dimensionality assessment in bifactor structures with multiple general factors: A network psychometrics approach. *Psychological Methods*. https://doi.org/doi.org/10.1037/met0000590.supp

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.

Kan, K.-J., Psychogyiopoulos, A., Groot, L. J., Jonge, H. de, & Ten Hove, D. (2024). Why do bi-factor models outperform higher-order g factor models? A network perspective. *Journal of Intelligence*, *12*(2), 18. https://doi.org/10.3390/jintelligence12020018

Kassambara, A. (2018). *Ggpubr: 'ggplot2' based publication ready plots*. Retrieved from https://CRAN.R-project.org/package=ggpubr

Lafontaine, M.-F., Brassard, A., Lussier, Y., Valois, P., Shaver, P. R., & Johnson, S. M. (2015). Selecting the best items for a short-form of the experiences in close relationships questionnaire. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000243

Lai, K. (2019). Creating misspecified models in moment structure analysis. *Psychometrika*, *84*(3), 781–801. https://doi.org/10.1007/s11336-018-09655-0

Laub, A. J. (2005). *Kronecker products* (Vol. 91). SIAM: Society for Industrial; Applied Mathematics.

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

Lorenzo-Seva, U., & Ferrando, P. J. (2019). A general approach for fitting pure exploratory bifactor models. *Multivariate Behavioral Research*, *54*(1), 15–30. https://doi.org/10.1080/00273171.2018.1484339

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

Maertens, R., Götz, F. M., Golino, H. F., Roozenbeek, J., Schneider, C. R., Kyrychenko, Y., et al.others. (2023). The misinformation susceptibility test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods*, 1–37. https://doi.org/10.3758/s13428-023-02124-2

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing fit in structural equation models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 389–402. https://doi.org/10.1080/10705511.2017.1389611

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

Montoya, A. K., & Edwards, M. C. (2021). The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. *Educational and Psychological Measurement*, *81*(3), 413–440. https://doi.org/10.1177/0013164420942899

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via monte carlo simulations. *Journal of Intelligence*, *3*(1), 2–20.

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

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407–422.

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28–56.

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

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/

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

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5–26.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137.

Rosseel, Y. (2012a). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y. (2012b). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

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

Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, *66*(2), 201–223.

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.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.

Shi, D., DiStefano, C., McDaniel, H. L., & Jiang, Z. (2018). Examining chi-square test statistics under conditions of large model size and ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 924–945.

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

Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 1–15. https://doi.org/10.1080/10705511.2019.1611434

Snyder, H. R., Young, J. F., & Hankin, B. L. (2017). Strong homotypic continuity in common psychopathology-, internalizing-, and externalizing-specific factors over time in adolescents. *Clinical Psychological Science*, *5*(1), 98–110.

Song, J., & Belin, T. R. (2008). Choosing an appropriate number of factors in factor analysis with incomplete data. *Computational Statistics & Data Analysis*, *52*(7), 3560–3569.

Steiger, J. H. (1980). Statistically based tests for the number of common factors. *The Annual Meeting of the Psychometric Society. Iowa City, IA. 1980.*

Tian, C., & Liu, Y. (2021). A rotation criterion that encourages a hierarchical factor structure. In M. Wiberg, D. Molenaar, J. González, U. Böckenholt, & J.-S. Kim (Eds.), *Quantitative Psychology* (pp. 1–8). Springer International Publishing. https://doi.org/10.1007/978-3-030-74772-5_1

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*(4), 421–459. https://doi.org/10.1007/BF02290601

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

Von Neumann, J. (1927). Wahrscheinlichkeitstheoretischer aufbau der quantenmechanik. *Nachrichten von Der Gesellschaft Der Wissenschaften Zu Göttingen, Mathematisch-Physikalische Klasse*, *1927*, 245–272.

Waller, N. G. (2018). Direct Schmid–Leiman transformations and rank-deficient loadings matrices. *Psychometrika*, *83*(4), 858–870. https://doi.org/10.1007/s11336-017-9599-0

Watanabe, H. (2000). A new view of the formal entropy as a measure of interdependence and its application to pattern recognition. *SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics: "Cybernetics Evolving to Systems, Humans, Organizations, and Their Complex Interactions"*, *4*, 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. *"2001 IEEE International Conference on Systems, Man and Cybernetics": "E-Systems and e-Man for Cybernetics in Cyberspace"*, *4*, 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

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

Xia, Y., & Yang, Y. (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

Yeung, R. W. (2008). *Information theory and network coding.* Springer Science & Business Media.

Yuan, K.-H., & Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology*, *56*(1), 93–110. https://doi.org/10.1348/000711003321645368