

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: www.tandfonline.com/journals/lssp20

# Mild-restricted confirmatory factor analysis: a new approach to assessing population factor models

Urbano Lorenzo-Seva

**To cite this article:** Urbano Lorenzo-Seva (2026) Mild-restricted confirmatory factor analysis: a new approach to assessing population factor models, Communications in Statistics - Simulation and Computation, 55:5, 1656-1674, DOI: 10.1080/03610918.2025.2453473

**To link to this article:** https://doi.org/10.1080/03610918.2025.2453473

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC

Published online: 20 Jan 2025.

Submit your article to this journal

Article views: 2093

View related articles

View Crossmark data

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lssp20

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
2026, VOL. 55, NO. 5, 1656–1674
https://doi.org/10.1080/03610918.2025.2453473

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

# Mild-restricted confirmatory factor analysis: a new approach to assessing population factor models

Urbano Lorenzo-Seva

Facultat de Ciències de l'Educació i Psicologia, Universitat Rovira i Virgili, Tarragona, Spain

**ABSTRACT**

Mild-restricted confirmatory factor analysis is a new method for computing confirmatory factor analysis (CFA). In CFA, each variable in the model is usually expected to be a pure indicator of a factor at the population level. However, even if a variable is strongly related to a single factor, it often cannot be considered a pure indicator because the other loading values of the variable are not exactly zero, although they are close to zero. This kind of variable can be said to be a close indicator of a factor. The CFA based on mild-restricted factor analysis helps to assess whether a factor model can be expected to exist at population level even if the variables are close indicators or none of the inter-factor correlations or the correlated errors are exactly zero.

**ARTICLE HISTORY**
Received 21 March 2024
Accepted 9 January 2025

**KEYWORDS**
Confirmatory factor analysis; Exploratory factor analysis; Factor model assessment; Restricted factor analysis; Unrestricted factor analysis

## 1. Introduction

Common factor analysis is the most popular method for identifying and studying the unobserved variables (i.e. factors or dimensions) that underlie a set of observed variables. While it was initially used to study general intelligence, it was soon extended to other domains like personality. Nowadays, it is still the most used statistical tool for examining the internal structure of scales and questionnaires in psychological and educational assessment.

The literature and applied research frequently differentiate between Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). In EFA, the analysis explores which factor model best suits the dataset to be analyzed. In CFA, the analysis assesses whether a factor model (probably determined in a previous EFA) can be expected to hold at the population level. While factor analysis was conceived as a pure exploratory technic, Jöreskog in 1969 proposed the bases for the confirmatory step. He proposed (1) to specify a priori as many parameters of the factor solution as possible; (2) to fit the restricted pattern; and (3) to assess the appropriateness of the factor solution with a proper test of the fit indices.

However, as Bollen pointed out in 1989, the distinction between EFA and CFA is more conceptual than model dependent. For example, an applied study using EFA in which the number of factors and the approximate structure are hypothesized in advance is more confirmatory than exploratory; on the other hand, an applied study using CFA that assesses different factor models modified 'ad hoc' to improve a poor fitting of the initially proposed model is more exploratory than confirmatory (Bollen 1989). EFA and CFA are not really different models, rather, as part of a common model, they are types of solutions that differ in the structural constraints they impose. As Ferrando (2021) proposes, if the artificial EFA vs CFA distinction is abandoned, then the

CONTACT Urbano Lorenzo-Seva urbano.lorenzo@urv.cat Facultat de Ciències de l'Educació i Psicologia, Universitat Rovira i Virgili, Ctra. de Valls s/n, Tarragona, Spain.

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

researcher will find a continuum of possibilities ranging from a totally unrestricted solution to a fully restricted solution. A recent review of this continuum can be found in Nájera, Abad, and Sorrel (2023).

The advice to shift the focus from the ambiguous EFA-CFA duality to the strength of the restrictions proposed during the analysis is not new (see, for example, Ferrando and Lorenzo-Seva 2000). An unrestricted solution does not restrict the factor solution, so unrestricted solutions can be obtained by a rotation of an arbitrary orthogonal solution. For the same data, all the unrestricted solutions will yield the same fit. A restricted solution, on the other hand, imposes restrictions on the whole factor space and cannot be obtained by a rotation of an unrestricted solution (Jöreskog 1969). While restricted confirmatory factor analysis is more determinate, and potentially more stable, it is also less flexible. This means that misspecifications in the model (for example, substantial cross-loading values are proposed to be zero in the factor model) can lead to biases in the parameters of the factor model that are estimated.

In the present document we propose a new methodological approach to compute confirmatory factor analysis. The aim is to avoid biases of the estimated parameters due to the misspecifications in the proposed factor model. As we explain below, the key idea is to estimate free parameters before the restrictions are applied. We name our approach *mild-restricted confirmatory factor analysis*.

### 1.2. Exploratory versus confirmatory factor analysis

In the classic factor analysis model, the correlation matrix between $m$ items in the population is decomposed as

$$\mathbf{\Sigma} = \mathbf{\Lambda\Theta\Lambda}' + \mathbf{\Psi}^2, \tag{1}$$

where $\mathbf{\Sigma}$ is the $m \times m$ inter-item correlation matrix, $\mathbf{\Lambda}$ is the $m \times r$ rotated pattern, $\mathbf{\Theta}$ is the $r \times r$, inter-factor correlation matrix, and $\mathbf{\Psi}$ is the $m \times m$ diagonal matrix containing the item residual standard deviations. In the model shown in expression (1) the residuals are all uncorrelated. For the particular case of $r = 1$, $\mathbf{\Lambda}$ becomes a vector, and $\mathbf{\Theta}$ a scalar of value 1. This method can be extended to allow residuals to be correlated. Now the correlation matrix $\mathbf{\Sigma}$ is decomposed as

$$\mathbf{\Sigma} = \mathbf{\Lambda\Theta\Lambda}' + \mathbf{\Psi\Sigma}_{\mathbf{uu}}\mathbf{\Psi}, \tag{2}$$

where $\mathbf{\Sigma}_{\mathbf{uu}}$ is the $m \times m$ residual correlation matrix. When the population model in expression (2) is estimated from a sample, it can be expressed as

$$\mathbf{R} = \mathbf{P\Phi\ P}' + \mathbf{UR}_{uu}\mathbf{U}. \tag{3}$$

The only additional assumption that is required for estimating (2) from (3) is that the sample inter-item correlation matrix $\mathbf{R}$ is positive definite (see Lorenzo-Seva and Ferrando 2021). The values observed in the sample model (i.e. expression 2) are expected to be unbiased estimates of the true values in the population model (i.e. expression 3).

#### 1.2.1. Exploratory factor analysis (EFA)

If the researcher does not have a hypothesis about any of the parameter values that define the model (2), then the decomposition in (3) is estimated using exploratory factor analysis: this means that unrestricted factor techniques are used. For example, the number of factors $r$ can be estimated using an objective method such as Parallel Analysis (Timmerman and Lorenzo-Seva 2011), the value of the loading values in $\mathbf{P}$ and the inter-factor correlation values can be estimated using Promin rotation (Lorenzo-Seva and Ferrando 2019), and the residual correlations can be estimated using Morgana (Ferrando, Hernandez-Dorado, and Lorenzo-Seva 2024).

If the researcher has a hypothesis about some of the parameter values in expression (2), exploratory factor analysis can still be used to compute a factor analysis that has a confirmatory aim. For example, if an $r$ number of factors is expected, and the first item is expected to be a close indicator of a particular factor, then the loading value of this item on such a factor is expected to be different from zero, while the other loading values of the item are expected to be close to zero. To assess this kind of hypothesis, oblique rotation to a partially specified target rotation can be used (Browne 1972) to estimate $\mathbf{P}$ and $\mathbf{\Phi}$ in expression (3). In addition, when the hypothesis is not a strong one, the proposed target can be objectively refined during the analysis (Lorenzo-Seva and Ferrando 2020).

### 1.2.2. Confirmatory factor analysis (CFA)

When the researcher has a strong hypothesis about the value of the parameters that define model (2), the decomposition in (3) is estimated using confirmatory factor analysis: this means that restricted confirmatory factor analysis techniques are used. The aim of the analysis is to assess whether a hypothesized factor model can be expected at the population level.

In confirmatory factor analysis, the corresponding scalar elements in expression (2) that are hypothesized to show a particular value are taken as fixed values in expression (3). For example, if an $r$ number of factors is proposed, and the first item is expected to be a pure indicator of a particular factor, then the loading value of this item on such a factor is proposed to be different from zero, while the other loading values of the item are proposed to be zero. To assess this hypothesis, the loading values proposed to be zero are fixed to be so in matrix $\mathbf{P}$, and only the loading value related to the first factor needs to be estimated. In the overall expression (3), the hypothesis can be a relaxed one (i.e. only a few parameter values are hypothesized), or a very strict one in which a large number of parameter values in $\mathbf{P}$, $\mathbf{\Phi}$, and $\mathbf{R}_{uu}$ are taken as fixed values (Ferrando and Lorenzo-Seva 2000).

It must be pointed out that the number of parameters that needs to be estimated is limited by the number of degrees of freedom available. The final degrees of freedom are

$$df = \frac{m(m-1)}{2} - f, \tag{4}$$

where $df$ are the degrees of freedom and $f$ is the number parameters that are taken as free values (i.e. they need to be estimated) in expression (3). On the other hand, to be able to identify the factor model, at least $r$-1 scalar elements of each column of $\mathbf{P}$ need to be specified. Finally, diagonal elements of $\mathbf{\Phi}$ are constrained to be 1.

Different computational approaches can be used depending on the assumptions. Li (2016) carried out an extensive simulation study to compare two different estimation methods (Robust Maximum Likelihood, and Diagonally Weighted Least Squares).

In this manuscript, we refer to confirmatory factor analysis as it is usually estimated as restricted confirmatory factor analysis (R-CFA) in order to clearly differentiate this traditional approach from our new proposal.

### 1.2.3. Controversy between EFA and CFA

In exploratory factor analysis, parameter values in expression (3) need to be estimated, and they are then taken as estimates of the corresponding parameter values at the population level (i.e. the parameter values in expression 2). In confirmatory factor analysis, only the parameter values in expression (3) that are not fixed need to be estimated. To do so, a confirmatory factor analysis first imposes the fixed scalar values in expression (3), and then estimates the values of the free ones.

The limitation of restricted factor analysis is that the misspecification of the expected true population value of a parameter in expression (2) might have an effect on the estimated values of the free parameters in expression (3). For example, let us suppose that the first item is hypothesized to be a pure indicator of factor one (i.e. all the secondary loading values are proposed to be exactly zero), while this item just happens to be a close indicator of factor one (i.e. all the secondary loading values are close to zero). As the item is just a close indicator, secondary loading values are not exactly zero (either at population level or at sample level). However, as the (erroneous) hypothesis proposes that the item is a pure indicator of the first factor, the secondary loadings in expression (3) are fixed to be exactly zero. The consequence is that the amount of variance accounted for by the secondary loading values at the population level is artificially displaced toward the estimated value of the first loading of the item in expression (3). If this kind of misspecification is carried out in a large number of items, then the amount of variance artificially displaced in the model will be large, and the estimates of the scalar values in $\mathbf{\Phi}$ and $\mathbf{R}_{uu}$ can also be overestimated. To summarize, the misspecification of the fixed parameters also affects the estimated values for the free parameters of the model in expression (3).

It must be said that the hypothesis proposed in the example of the first item is not completely wrong: it is just partially wrong (because the item is actually an indicator of the first factor). The value of the goodness-of-fit indices should, to some extent, detect that the proposed hypothesis is partially correct. However, as the fixed loading values are misspecified and, in consequence, the value of the free loading is overestimated in expression (3) with respect to the true value in expression (2), the goodness-of-fit indices will suggest a very bad fit for the proposed hypothesis.

While confirmatory factor analysis is nowadays considered as a less suitable analysis technique (Ferrando and Lorenzo-Seva 2000; Steenkamp and Maydeu-Olivares 2023), there are still situations in which researchers might need to assess factor models in which some parameters have been fixed to show arbitrary values at the population level. The fixed values can be an exact zero (for example, factors are zero correlated), or values already observed in previous samples or studies.

### 1.2.4. Exploratory Structural Equation Modeling (ESEM)

To solve the controversy between EFA and CFA, Asparouhov and Muthén (2009) focused on the fact that the two approaches are two extremes of the continuum between the pure exploratory aim, at one end, and the pure confirmatory aim, at the other end. Therefore, they proposed the new label *Exploratory Structural Equation Modeling* (ESEM). The main idea is that (1) any factor analysis (exploratory or confirmatory) is a structural equation model, (2) standard error coefficients should be provided, and (3) the goodness-of-fit of the model must be assessed using the same indices as the ones typically used in CFA.

If ESEM is applied with a full exploratory aim, a rotation criterion must be specified. For example, if Quartimin rotation (Jennrich and Sampson 1966) is specified, then the method applied in practice would be ESEM-Quartimin. If the ESEM is applied with a less exploratory aim (for example, because a hypothesis is available), then ESEM-Target can be used. The ESEM-Target model is constructed by setting all loadings that are not significantly different from 0 as targets and the rotation criteria is the partially specified target rotation (Browne 1972). In ESEM-Target, the targets specified are not fixed zero values in the rotated pattern matrix (as is the case in CFA): zero target values can become large in the rotated pattern matrix if a close to zero value is not admissible for the analyzed data at hand.

A step closer to the CFA would be the Independent-Clusters-Basis (ICB) model proposed by McDonald (1999, p 181). In an ICB model a number of items per factor are identified as pure indicators of each factor (i.e. non-salient value loadings of each pure indicator are restricted to be exactly zero in the pattern matrix). In the context of ESEM, the model estimated would be

ESEM-ICB. In ESEM-ICB at least one pure indicator per factor is required, and it is advisable to define up to three pure indicators per factor.

Finally, a completely confirmatory aim could be applied: then all the items would be considered as a pure indicator of a factor, and all non-salient loadings would be restricted to be exactly zero. In this case, the model estimated would be SEM-CFA.

## 2. A toy example based on a real dataset

To visualize the differences between the approaches already presented in the introduction and our new proposal, we propose to analyze a toy dataset. It is an artificial dataset constructed with Monte Carlo data simulation techniques and is inspired by a real dataset by Stankov, Roberts and Spilsbury (1994). These authors analyzed a correlation matrix of 22 variables by factor analysis, and three factors were identified: crystallized intelligence (Gc), fluid intelligence (Gf) and short-term acquisition and retrieval function (SAR). This example is useful because each factor was expected to be defined by a number of pure indicator variables: 8 variables define Gc, 9 variables define Gf, and 5 variables define SAR. However, few of the expected non-salient loadings were really close to zero, and the three factors show a moderate correlation. For the population factor solution, we used the factor solution reported by the researchers to produce the scores of 5,000 participants in the 22 variables.

The correlation matrix between the 22 variables was factor analyzed to extract three factors, and the solution was analyzed with different methods in the Exploratory vs Confirmatory continuum. We computed ESEM-Quartimin, ESEM-Target, and ESEM-ICB. In ESEM-Target, all the loading values that were not expected to be salient loadings were defined as zero values in the target. In ESEM-ICB, only six variables were defined as pure indicators (two per factor). It must be pointed out that the analysis of the same data using different methods should not be used in real data analysis: each analysis should be carried out with a different sample (all of them representative of the population). However, for illustrative purposes, it is useful to analyze the same data with different methods. As the data analyzed with each method is always the same, the numerical outcomes reported by each method should be quite similar, and the conclusions should be the same after each analysis.

Table 1 shows that the three methods produced similar numerical outputs for the estimates of the salient loading values, and that these values agree with the expected values. Only one variable seems doubtful: *Letter Counting*. As we only defined six indicator variables (two for each factor) in ESEM-ICB, it can be seen that the minimum restrictions were applied to the pattern matrix: 12 loading values were restricted to be exactly zero, while 54 loading values were estimated. Table 2 shows the corresponding inter-factor correlation matrices obtained with each ESEM method.

Table 2 shows that ESEM-Quartimin and ESEM-Target led to the same correlation values, while ESEM-ICB reported some differences: one correlation was larger than the ones reported by the other two methods, and two correlations were slightly lower. Table 3 summarizes the goodness-of-fit indices. Root Mean Square Error of Approximation (RMSEA) assesses the fit related to the complexity of the model; the Comparative Fit Index (CFI) describes the comparative fit of the proposed solution with respect to the null model of independence; and the Goodness of Fit Index (GFI) indicates how closely a factor model matches the observed data.

ESEM-Quartimin and ESEM-Target obtained the same goodness-of-fit indices. ESEM-ICB had a questionable value in relation to the RMSEA. To an applied researcher, the outcomes of these analyses suggest that each factor (Gc, Gf, and SAR) is properly defined by the expected indicator variables.

A consequence of this conclusion is that, to compute the participants' scores in each factor, only the variables with a salient loading in each factor need to be considered (for a further discussion on this subject, see Ferrando and Lorenzo-Seva 2021). For example, to obtain scores in

**Table 1.** Patterns obtained by the three ESEM methods related to the toy example.

| Variables | ESEM-Quartimin | | | ESEM-Target | | | ESEM-ICB | | |
|---|---|---|---|---|---|---|---|---|---|
| | Gc | Cf | SAR | Gc | Cf | SAR | Gc | Cf | SAR |
| Information | **.91** | −0.13 | .01 | **.96** | −0.24 | −0.05 | **.92** | −0.22 | .03 |
| Digit Span | .03 | .22 | **.61** | .02 | .11 | **.65** | – | – | **.69** |
| Vocabulary | **.92** | −0.08 | −0.01 | **.98** | −0.19 | −0.07 | **.94** | −0.17 | .01 |
| Arithmetic | **.48** | .12 | .19 | **.50** | .03 | .17 | **.48** | −0.01 | .24 |
| Comprehension | **.87** | .04 | −0.02 | **.91** | −0.06 | −0.08 | **.85** | – | – |
| Similarities | **.64** | .11 | −0.03 | **.67** | .04 | −0.07 | **.67** | – | – |
| Picture Completion | .30 | **.49** | .20 | .30 | **.42** | .19 | .27 | **.36** | .28 |
| Picture Arrangement | .07 | **.60** | .08 | .05 | **.58** | .09 | .04 | **.53** | .15 |
| Block Design | −0.04 | **.78** | −0.05 | −0.06 | **.80** | −0.05 | – | **.74** | – |
| Object Assembly | .08 | **.83** | −0.06 | .06 | **.84** | −0.07 | – | **.85** | – |
| Digit Symbol | −0.22 | **.69** | .00 | −0.26 | **.72** | .01 | −0.25 | **.68** | .05 |
| General Information | **.58** | .21 | −0.05 | **.60** | .15 | −0.09 | **.57** | .15 | −0.01 |
| Letter Series | −0.10 | **.75** | .15 | −0.14 | **.74** | .16 | −0.14 | **.67** | .23 |
| Concealed Words | −0.03 | **.30** | −0.08 | −0.04 | **.32** | −0.08 | −0.04 | **.32** | −0.07 |
| Digit Span Forward | .02 | .21 | **.39** | .02 | .14 | **.41** | – | – | **.50** |
| Digit Span Backward | .00 | .14 | **.20** | .00 | .10 | **.22** | −0.01 | .06 | **.24** |
| Esoteric Analogies | **.54** | .28 | .16 | **.56** | .20 | .13 | **.52** | .16 | .22 |
| Cattell's Matrices | .19 | **.53** | −0.08 | .18 | **.53** | −0.10 | .17 | **.53** | −0.04 |
| Word Associations | .14 | **.40** | −0.26 | .14 | **.43** | −0.28 | .14 | **.47** | −0.26 |
| Letter Counting | −0.01 | .33 | **.24** | −0.02 | .29 | **.26** | −0.03 | **.22** | .30 |
| Synonym Vocabulary | **.55** | .11 | .20 | **.57** | .01 | .18 | .54 | −0.03 | **.26** |
| Crowder's Memory Span | .01 | −0.07 | **.98** | .02 | −0.20 | **1.03** | −0.03 | −0.46 | **1.14** |

The loading values that are expected to be salient are written in bold.

**Table 2.** Inter-factor correlation matrices obtained by the three ESEM methods related to the toy example.

| | ESEM-Quartimin | | | ESEM-Target | | | ESEM-ICB | | |
|---|---|---|---|---|---|---|---|---|---|
| | Gc | Cf | SAR | Gc | Cf | SAR | Gc | Cf | SAR |
| Gc | 1 | | | 1 | | | 1 | | |
| Cf | .26 | 1 | | .34 | 1 | | .26 | 1 | |
| SAR | .17 | .30 | 1 | .41 | .41 | 1 | .37 | .52 | 1 |

**Table 3.** Goodness-of-fit indices of the ESEM methods related to the toy example.

| Method | RMSEA | CFI | GFI |
|---|---|---|---|
| ESEM-Quartimin | .03 | .99 | .99 |
| ESEM-Target | .03 | .99 | .99 |
| ESEM-ICB | .09 | .99 | .99 |

Gc, only the scores in the eight variables that show a salient loading in factor Gc would need to be considered. However, when factor scores are calculated only taking salient loadings into account, it has been assumed implicitly that the values of non-salient loadings in this factor are so close to zero that they can be considered as zero values. The applied researcher would then aim to assess whether these secondary loadings can actually be considered zero, and would compute a restricted CFA in order to assess this. We computed such a CFA for the dataset, with the expected non-salient loadings restricted to the value of zero. As we are computing the analysis based on the same data as in the previous analyses, we expected that the values estimated should be similar to the ones obtained in the previous analyses. Just remember that in a real research situation, a new sample should be used for this new analysis, and that our aim here is just a pedagogical one to check whether the expectations are fulfilled. The outcome is shown in Table 4 (labeled as R-CFA).

The applied researcher would be surprised to discover that estimated values in R-CFA frequently differ from the ones obtained in the three previous analyses. For example, variable

**Table 4.** Patterns obtained by the two CFA methods related to the toy example.

| Variables | R-CFA | | | MR-CFA | | |
|---|---|---|---|---|---|---|
| | Gc | Cf | SAR | Gc | Cf | SAR |
| Information | .76 | – | – | .84 | – | – |
| Digit Span | – | – | .74 | – | – | .48 |
| Vocabulary | .79 | – | – | 1.01 | – | – |
| Arithmetic | .63 | – | – | .40 | – | – |
| Comprehension | .82 | – | – | .96 | – | – |
| Similarities | .65 | – | – | .68 | – | – |
| Picture Completion | – | .78 | – | – | .40 | – |
| Picture Arrangement | – | .66 | – | – | .58 | – |
| Block Design | – | .68 | – | – | .79 | – |
| Object Assembly | – | .80 | – | – | .83 | – |
| Digit Symbol | – | .50 | – | – | .73 | – |
| General Information | .64 | – | – | .61 | – | – |
| Letter Series | – | .72 | – | – | .73 | – |
| Concealed Words | – | .22 | – | – | .34 | – |
| Digit Span Forward | – | – | .53 | – | – | .62 |
| Digit Span Backward | – | – | .30 | – | – | .24 |
| Esoteric Analogies | .77 | – | – | .56 | – | – |
| Cattell's Matrices | – | .58 | – | – | .56 | – |
| Word Associations | – | .32 | – | – | .42 | – |
| Letter Counting | – | – | .50 | – | – | .28 |
| Synonym Vocabulary | .70 | – | – | .59 | – | – |
| Crowder's Memory Span | – | – | .67 | – | – | .92 |

**Table 5.** Inter-factor correlation matrices obtained by the CFA methods related to the toy example.

| R-CFA | | | MR-CFA | | |
|---|---|---|---|---|---|
| Gc | Cf | SAR | Gc | Cf | SAR |
| 1 | | | 1 | | |
| .37 | 1 | | .31 | 1 | |
| .46 | .55 | 1 | .44 | .50 | 1 |

*Picture Completion*, that in the previous analyses had loading values in the range of 0.36 to 0.49, in R-CFA showed a value of .78. However, the overestimation was not systematic: the variable *Crowder's Memory Span*, that in the previous analyses had loading values in the range of 0.98 to 1.14, in R-CFA showed a value of .67. Table 5 shows the inter-factor correlation matrix. As can be observed, the inter-factor correlations reported by R-FCA were systematically larger than the ones obtained with the ESEM methods. These misleading differences in the outcome are not new and are reported in the literature (see, for example, Ferrando and Lorenzo-Seva 2000; Asparouhov and Muthén 2009). These differences are also the reason why R-CFA is not often used in applied research.

The goodness-of fit indices related to the R-CFA analysis were $RMSEA = 0.11$, $CFI = 0.91$ and $GFI = .94$. This outcome means that a factor model that was acceptable from the ESEM point of view, is not acceptable from the CFA point of view. The reason for the unacceptable values reported by *RMSEA* and *CFI* is that loading restrictions in line with confirmatory factor analysis (i.e. non-salient loadings equal to zero) lead to a distortion and misestimation of the parameters of the model.

The question that we face in this manuscript is whether it is possible to estimate the parameters of the model in a CFA when a large number of parameters are restricted to being zero. In R-CFA, the free parameters are estimated after the zero constraints have been imposed. Our proposal is to reverse the order: to first estimate the free parameters, and then to impose the zero constraints. Our approach is still a restricted CFA, although it is applied cautiously, so we call it a Mild Restricted Confirmatory Factor Analysis (MR-CFA). We will explain the approach

in detail in the next section, but now we will look at the outcome that MR-CFA produces with the toy example.

The model that we tested with MR-CFA is the same as with the R-CFA. The pattern matrix is shown in Table 4, and the inter-factor correlation matrix is shown in Table 5. If we compare the values estimated with MR-CFA, they are more similar to the ones obtained with the ESEM approaches. To visualize the similitudes, we computed the Mean Root Mean Square (see expression 8), between the salient loading values among ESEM-Quartimin, ESEM-Target, ESEM-ICB, R-CFA and MR-CFA. Then we analyzed the matrix of distances using multidimensional scaling. The same analysis was performed for the inter-factor correlation matrix. The corresponding scatterplot is presented in Figure 1. The closer the points in the graph, the more similar the salient loading values are. As can be seen in the figure, ESEM approaches are quite similar among them, and R-CFA is far away from them. MR-CFA is closer to the ESEM approaches than to R-CFA.

The configuration observed for the inter-factor correlations (Figure 2) shows how the main dimension arranges the methods from the most EFA (i.e. ESEM-Quartimin) to the most CFA methods. MR-CFA seems to be in a close position between ESEM-Target and ESEM-ICB on one side, and R-CFA on the other side.

Finally, the goodness-of-fit indices related to MR-CFA had the values $RMSEA = 0.05$, $CFI = 0.98$ and $GFI = .96$. The values of these indices, even if they were slightly worse than the corresponding values obtained in the ESEM approaches, were comparable to them. The conclusion is that the hypothesized model can be expected to be found at population level.

In the next section, we explain in detail this new confirmatory factor analysis that we name Mild-Restricted Confirmatory Factor Analysis (MR-CFA). The usefulness of this new proposal is tested in a simulation study and illustrated by the analysis of a real dataset.

## 3. Mild-restricted confirmatory factor analysis (MR-CFA)

In mild-restricted confirmatory factor analysis (MR-CFA) we propose to estimate the value of the free parameters first, and then to impose this value on the restricted parameters. If the hypothesized factor model is perfectly correct at the population level, then the fit observed at the sample level should be similar to the fit observed in a restricted confirmatory factor analysis. However, if the hypothesized factor model is only partially correct, as the free parameters are estimated without the constraint of the fixed values, we expect the estimated values of the free parameters to be more accurate than in a restricted confirmatory factor analysis.

MR-CFA is computed in three steps: first, the free parameters are estimated using unrestricted factor techniques; second, the restrictions proposed by the model that is being tested are applied; and third, the goodness-of-fit indices based on the LOSEFER chi-square transformation are assessed (for a detailed revision of LOSEFER, please see Lorenzo-Seva and Ferrando 2023).

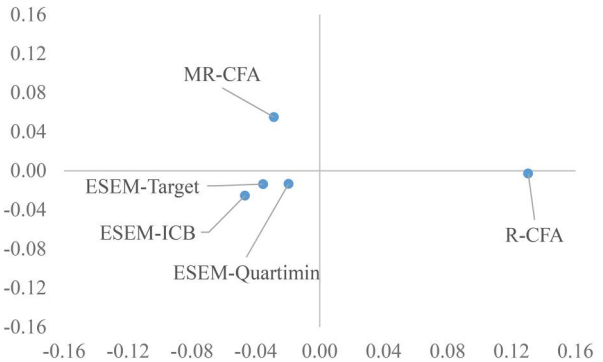



**Figure 1.** Scatterplot of distances related to salient loading values among methods.

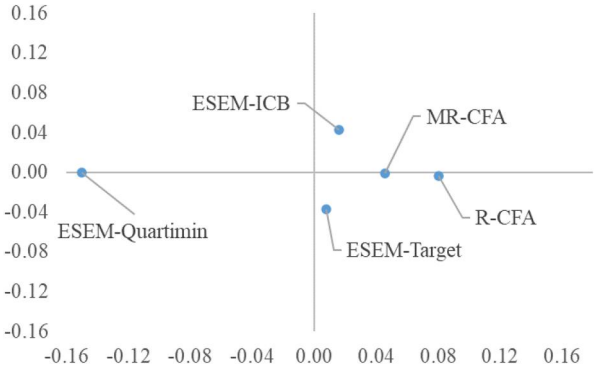



**Figure 2.** Scatterplot of distances related to inter-factor correlations among methods.

### 3.1. Estimating free parameters using unrestricted factor techniques

The first step treats the situation as a typical exploratory factor analysis in which all the scalar parameters of the model need to be estimated. The MORGANA factor analysis is used to extract factors (Ferrando, Hernandez-Dorado, and Lorenzo-Seva 2024). This extraction method is an EFA approach in which both the common factor solution (matrix **P**) and the residual matrix (matrix $\mathbf{R}_{uu}$) are unrestricted. MORGANA aims to overcome a frequent criticism of exploratory factor analysis (EFA), that is, that it does not allow correlated residuals to be estimated. MORGANA is an EFA approach in which both the common factor solution and the residual matrix are unrestricted (i.e. the correlated residuals do not need to be specified a priori). The estimation procedure is a two-stage approach. The first stage aims to determine and estimate the correlated errors that are the most substantial. The second stage consists essentially of fitting a conventional factor solution to a reduced correlation matrix in which the first-order correlations are partialled out using the residual structure determined in the first stage.

If $r = 1$, then the first step has finished. When $r > 2$, a partially specified target is defined. The specified elements of the target are the scalar elements for which the researcher has defined a specific value, while all other elements are unspecified in the target. The rotation applied is the oblique rotation to a partially specified target (Browne 1972): this is a Procrustes rotation in which the rotated loading values are intended to be as close as possible to their corresponding specified element in the target, but not constrained to be exactly equal to the value.

Please note that, as the rotation computed is an oblique one, the correlation between factors is not constrained to be zero. If in the population factor model the inter-factor correlation matrix is diagonal (i.e. the factors are independent), the observed inter-factor correlation matrix in the sample will be close to (but not exactly) an identity matrix.

At the end of this first step, matrices **P**, **Φ** and $\mathbf{R}_{uu}$ are available but still not in their final form.

### 3.2. Applying restrictions proposed by the model

Once all the values in **P**, **Φ** and $\mathbf{R}_{uu}$ have been estimated using exploratory factor analysis, the values hypothesized are now inserted in the corresponding positions in **P**, **Φ** and $\mathbf{R}_{uu}$. The proposed values can be either zero or an arbitrary value that the researcher hypothesizes to be the true value in the population.

By the end of this first step, the matrices **P**, **Φ** and $\mathbf{R}_{uu}$ are available. While some values are not restricted in the hypothesis, the actual values of the free parameters were estimated before the restrictions were imposed. For this reason, if the hypothesis is partially wrong, the estimated values of the free parameters can be expected not to be biased.

### 3.3. Assessing the goodness-of-fit indices based on the LOSEFER chi-square transformation

To assess whether the proposed hypothesis can be accepted at the population level, we need to inspect the goodness-of-fit index values. Some of the most popular indices (i.e. *RMSEA* and *CFI*) are computed using the chi-square statistic derived from the minimum fit function value adjusted using the chi-square statistic derived from the minimum fit function value. In addition, *GFI* can also be derived from *RMSEA* (Maydeu-Olivares, Ximénez, and Revuelta 2024). LOSEFER is an intensive computing approach that uses non-linear transformations to bring the distribution of the resulting fit statistic closer to the expected chi-square distribution. As LOSEFER is based on Monte Carlo simulations, the analysis that has been applied to the sample data must also be applied to each artificial dataset. LOSEFER is based on previous developments, particularly the bootstrapped approach by Bollen and Stine (1992): the difference is that with LOSEFER, to estimate the distribution of the chi-square when the null-hypothesis is true, the pseudo-chi-squared distribution obtained using Monte Carlo simulations is transformed to fit the four moments of the theoretically expected chi-square distribution. In this way, LOSEFER is more accurate than the Bollen and Stine approach.

## 4. Simulation study

### 4.1. Factor model generation

The simulated data were generated using a linear common factor model. The common factor model at the population level was based on the following characteristics:

1. Ten observed items were indicators of two common factors (five items for each factor).
2. The salient loading values were drawn uniformly from the range [.35, .75]. For the secondary loading values two conditions were considered: (1) secondary loadings equal to zero; and (2) secondary loadings drawn uniformly from the range [.05, .30]. In the first condition, items were pure indicators of a factor. In the second condition, items were close indicators of a factor. The sign of each loading value was drawn at random with equal probability for positive and negative values.
3. The two common factors were perfectly independent (i.e., correlation between factors equal to zero).
4. All correlated errors were exactly zero.

Monte Carlo simulation techniques were used to generate samples with $N = 1{,}000$ from the population solution as follows. The population loading matrix $\mathbf{A}(r \times m)$ had a constant number of factors $r = 2$ and variables $m = 10$. Once $\mathbf{A}$ was available, the population correlation matrix $\mathbf{R}$ was obtained as,

$$\mathbf{R} = \mathbf{A}\mathbf{A}^{'} + \mathbf{U}, \tag{5}$$

where $\mathbf{U}$ is a diagonal matrix of unicities. Let's decompose $\mathbf{R}$ using Cholesky's method,

$$\mathbf{L} = \text{chol}(\mathbf{R}), \tag{6}$$

and, if $\mathbf{Z}$ $(N \times m)$ is a random matrix with columns distributed normally $N(0, 1)$, the product

$$\mathbf{X} = \mathbf{Z}\mathbf{L} \tag{7}$$

produces a sample matrix $\mathbf{X}(N \times m)$ of observed scores. Matrix X was analyzed using MR-CFA and restricted confirmatory factor analysis (R-CFA) based on the hypothesis explained below. A total of 1,000 samples were generated.

### 4.2. Proposed hypothesis at the population level

The hypothesis proposed for the common factor model in the population was: (a) each item is a pure indicator of its corresponding factor in the population; (b) common factors correlate zero with each other; and (c) item correlated errors are zero. As can be observed, the hypothesis proposes that each item is a pure indicator of a factor. In the first condition related to secondary loadings, this hypothesis is completely true: when applied to this condition, we call this hypothesis a Completely True Hypothesis (CTH). In the second simulation condition related to secondary loading values, the population model is generated with items as close indicators (i.e. secondary loadings are close to zero, but not exactly zero). As the hypothesis is not perfectly consistent with the common factor model in the population, the analysis of any sample drawn from the population should be in partial agreement with the hypothesis, and the goodness-of-fit indices should report a reasonable fit. When applied to the second simulation condition, we called this hypothesis the Partially True Hypothesis (PTH). Figure 3 that helps to visualize the factor models at population level and the hypotheses tested.

Another hypothesis was proposed for the common factor model. This hypothesis is identical to the previous one, with the only exception being that it only relates four items (i.e. two items per factor) to the correct factor in the population model. As this hypothesis is mainly erroneous at population level, the goodness-of-fit indices should report an unacceptable fit of the sample factor solution to the proposed hypothesis at the population level. We labeled this hypothesis as the False Hypothesis (FH).

### 4.3. Factor analysis of sample data

For each data matrix **X**, a correlation matrix was computed. Factor analysis was also computed using two approaches.

1. Mild-restricted confirmatory factor analysis (MR-CFA). This was computed as discussed in the section above. The aim was to test the three hypotheses (CTH, PTH and FH). CTH was expected to show a perfect sample fit, PTH was expected to show an acceptable sample fit, and FH was expected to produce an unacceptable sample fit.

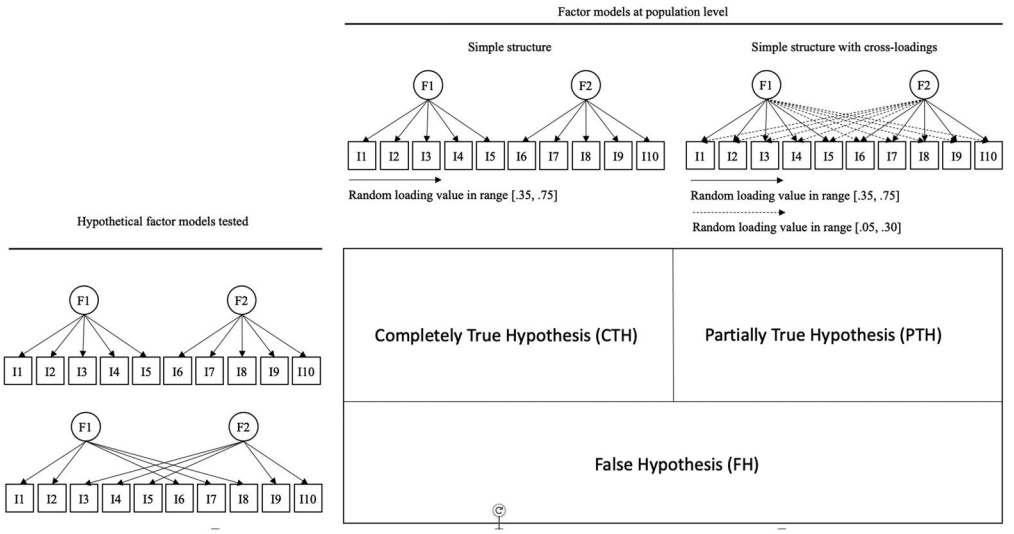



Figure 3. Factor models at population level and hypothetical factor models tested in the simulation study.

2. Restricted confirmatory factor analysis (R-CFA). To assess CTH and PTH in terms of a restricted factor model, the LISREL software was used to fit the sample factor model. Unweighted Least Squares was used as the extraction method.

### 4.4. Assessment of sample model fit to the proposed hypothesis

To assess the fit of the sample models, we computed the bias between the population and the sample-estimated loading matrices as the root of mean squared error,

$$bias = \sqrt{\frac{\sum \left(p_{ij} - \lambda_{ij}\right)^2}{K}}, \tag{8}$$

where $K$ is the number of samples generated. In addition, the goodness-of-fit indices were chi-square/*df*, *RMSEA* and *CFI*. Mean and standard deviations of *bias* and the goodness-of-fit indices were reported. To analyze the size of the effects among the differences of the means reported, we computed Cohen's $d$ (1998). This author proposed the following interpretation of the $d$ values: between 0.2 and 0.4, is a small effect; between 0.5 and 0.7, is a medium effect; and a value higher than 0.8 is a large effect.

We also counted the number of times the CTH and the PTH produced a value for the goodness-of-fit indices that would lead to the proposed hypothesis being rejected. As threshold values we used: chi-square/$df > 2$ (Tabachnick and Fidell 2007), $RMSEA > 0.07$ (Steiger 2007) and $CFI < 0.95$ (Hu and Bentler 1999). If any of these conditions were observed for the data at hand, we considered that the hypothesis was rejected. Finally, we also counted the number of times the FH produced a value for the goodness-of-fit indices that would lead to accept the proposed hypothesis. Now, we considered that the hypothesis was accepted if chi-square/$df \leq 2$, $RMSEA \leq 0.07$ and $CFI \geq 0.95$.

### 4.5. Results of the simulation study

Averages and standard deviations of the various indices obtained from the simulation study are shown in Table 6 and Figure 4 for the CTH and PTH. In the case of the CTH, while both approaches produced goodness-of-fit values that lead to the proposed model being accepted, MR-CFA seemed to slightly underestimate the true value of the loading values. Restricted confirmatory factor analysis produced the best model fit in terms of chi-square/*df* and *RMSEA*. None of the approaches produced any incorrect rejection of the hypothesis.

In the case of PTH, MR-CFA obtained the least biased solutions on average. In addition, the rate of overestimation of the true loading values (i.e. the proportion of estimated loading values that were larger than the corresponding population loading values) produced by MR-CFA was 47.1%, while the rate produced by R-CFA was 54.3%. In terms of the goodness-of-fit indices, MR-CFA reported the best fit of the sample data to the population model. At the same time, the average of the goodness-of-fit indices suggested that the fit was not perfect, which is to be expected because the hypothesis proposed is not perfectly true in the population. It should also

**Table 6.** Mean of goodness-of-fit indices for CTH and PTH. Standard deviations are given in parentheses.

| Hypothesis | Method | bias | chi-square/*df* | RMSEA | CFI | Rejection |
|---|---|---|---|---|---|---|
| CTH | MR-CFA | .037 | 1.14 (0.27) | .010 (.009) | .996 (.004) | 0% |
| | R-CFA | .031 | 1.00 (0.25) | .006 (.008) | .999 (.004) | 0% |
| PTH | MR-CFA | .042 | 2.07 (2.21) | .025 (.026) | .977 (.040) | 5.7% |
| | R-CFA | .048 | 10.25 (3.27) | .095 (.017) | .775 (.062) | 93.2% |

Rejection is decided when chi-square/$df > 2$, RMSEA $> 0.07$, or CFI $< 0.95$.

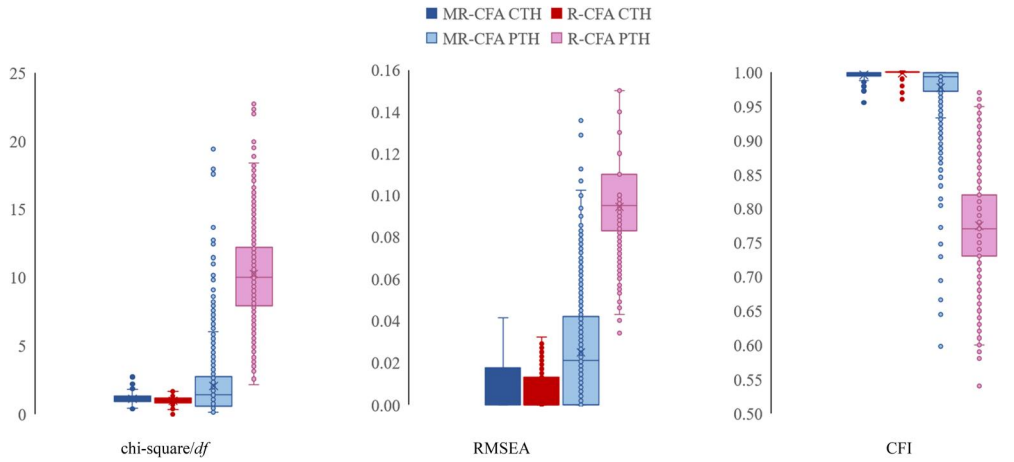



**Figure 4.** Graphical representation of goodness-of-fit indices related to the estimate methods in the different conditions of the simulation study. MR-CFA: Mild restricted confirmatory factor analysis. R-CFA: Restricted confirmatory factor analysis. CTH: Completely true hypothesis. PPH: Partially true hypothesis

be pointed out that restricted-factor analysis had a tendency to overestimate the loading values. Concerning the rejection of the proposed hypothesis, MR-CFA incorrectly rejected the hypothesis at a low rate. On the other hand, R-CFA nearly always rejected the hypothesis.

We computed Cohen's d to assess the effect size related to the differences in chi-square/*df* when MR-CFA analyzed the two conditions (CTH and PTH). The value was 0.681 (i.e. an intermediate effect): this effect size was congruent with the fact that in the second condition the hypothesized model was only partially wrong. However, in the case of R-CFA, Cohen's d was 4.147 (i.e. a very large effect): this effect size seems excessive. The conclusion is that, even when only a few errors were present in the hypothesized model, R-CFA suggested that the hypothesis was completely wrong.

As for the FH, the average ratio of chi-square/*df* was 34.82, and the goodness-of-fit indices consequently reported a bad fit of the model ($RMSEA = 0.162$ and $CFI = 0.556$). This means that MR-CFA detected that the hypothetical model proposed should be discarded at the population level. The percentage of times that the hypothesis was erroneously accepted was 7.75%. As the population model is a small one, and the size of the salient loading values is drawn from a wide range, some of the sample loading matrices could be, to some extent, close to the proposed hypothesis. For this reason, the percentage of times that the hypothesis was accepted seems reasonable.

## 5. Illustrative example

The Statistical Anxiety Test (Lorenzo-Seva, Vigil-Colet, and Ferrando 2022) was used for the illustrative example. All the items are rated on a 5-point graded format. They are designed to measure three factors related to anxiety: *Examination anxiety* (EX), *Asking for help anxiety* (AH), and *Interpretation anxiety* (IN). In addition, 5 items are designed to control response bias due to Social Desirability (SD). A sample of 682 undergraduate students from the first year of a bachelor's degree in psychology, between 18 and 60 years old (mean = 20.5, standard deviation =4.9), answered the test (Lorenzo-Seva, Vigil-Colet, and Ferrando 2024). They were informed that they participated in the study as an anonymous sample, and that no personal information was recorded.

The inter-item polychoric correlation matrix was found to be positive-definite, and showed a Kaiser-Meyer-Olkin (KMO) value of .863. The sample was divided into two subsamples using the Solomon method (Lorenzo-Seva 2022). The first subsample was used to compute an EFA using exploratory factor analysis. The aim of this first analysis was to explore whether there were any

Table 7. EREC indices related to the pair of items obtained in the first subsample.

| Pair | ITEM | EREC |
|---|---|---|
| 8 | I find having to interpret the meaning of a table in a journal article very distressing. | .281 |
| 12 | If I don't understand the statistical analyses described in a journal article, I keep calm and study them until I understand. | |

substantial correlated errors. The outcome of this first exploratory analysis was used to propose a hypothetical model. The second subsample was used to assess whether the hypothetical model proposed can be expected to be true in the population. This second analysis was a CFA, which compared MR-CFA and R-CFA. We used LISREL to compute R-CFA. In addition, we computed ESEM-ICB for comparative purposes.

### 5.1. Exploratory factor analysis

To assess whether there were any substantial correlated errors, the EREC index was computed (Ferrando, Hernandez-Dorado, and Lorenzo-Seva 2022). As shown in Table 7, the analysis detected a single pair of items for which errors are correlated. Both items share the term "journal article", so the correlation between errors could be explained by the content of items.

We used Morgana factor analysis to extract 4 factors. The value of the correlation between the errors of this pair of items was −0.272.

### 5.2. Hypothetical model proposed after the EFA in the first subsample

The outcome obtained in the first subsample and the factor solution obtained with this test in previous studies (see Lorenzo-Seva, et al. 2022) helped to propose the following model:

1. 15 items related to anxiety define three dimensions that are correlated.
2. Five items defined a dimension related to SD that is independent to anxiety dimensions.
3. The loadings of each item on their corresponding factor are defined as free parameters, while the other loadings are fixed to be zero.
4. The correlation between the three anxiety factors and the dimension of SD are set to correlate zero in the population.
5. The residual correlation between the pair of items observed in the first subsample is taken as a fixed value.

### 5.3. Mild-restricted confirmatory factor analysis (MR-CFA) of the second subsample

MR-CFA was computed as explained in the sections above to assess whether the model proposed after the first sample had been analyzed could be expected in the population. The sample analyzed was the second subsample. The loading matrix obtained after MR-CFA is shown in Table 8, and the inter-factor correlation matrix is shown in Table 9.

The goodness-of-fit indices suggest that the fit was appropriate: Chi-squared/$gl = 1.38$, $RMSEA = 0.032$, $CFI = 0.989$ and $GFI = .983$. The conclusion is that the hypothesized model can be expected to be found at population level.

### 5.4. Restricted confirmatory factor analysis (R-CFA) of the second subsample

Restricted factor analysis was computed using LISREL. The LISREL code can be found in the Appendix. Once again, the aim was to assess whether the model proposed after the first sample

**Table 8.** Loading matrix in the second subsample obtained with MR-CFA, ESEM-IC and R-CFA.

| | | MR-CFA | | | | ESEM-ICB | | | | R-CFA | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Items | | EX | AH | IN | DS | EX | AH | IN | DS | EX | AH | IN | DS |
| 1 | EX | .71 | – | – | – | .80 | – | – | – | **.86** | – | – | – |
| 2 | AH | – | **.87** | – | – | – | .86 | – | – | – | .86 | – | – |
| 3 | IN | – | – | **.57** | – | −0.11 | −0.12 | .51 | .08 | – | – | .54 | – |
| 4 | SD | – | – | – | .25 | .00 | −0.03 | .21 | .25 | – | – | – | **.38** |
| 5 | EX | .69 | – | – | – | −0.76 | – | – | – | **.79** | – | – | – |
| 6 | AH | – | **.90** | – | – | – | **.90** | – | – | – | .87 | – | – |
| 7 | IN | – | – | .46 | – | .22 | .22 | .38 | .13 | – | – | **.62** | – |
| 8 | SD | – | – | – | .30 | .10 | −0.08 | −0.34 | .28 | – | – | – | **.35** |
| 9 | EX+ | .85 | – | – | – | **.89** | −0.06 | .15 | .11 | .77 | – | – | – |
| 10 | AH | – | .76 | – | – | .09 | **.80** | .16 | −0.02 | – | .77 | – | – |
| 11 | IN | – | – | **.58** | – | −0.06 | −0.13 | .52 | .03 | – | – | .45 | – |
| 12 | SD | – | – | – | **.66** | – | – | – | .65 | – | – | – | .64 |
| 13 | EX | .52 | – | – | – | −0.59 | .12 | .06 | .03 | **.58** | – | – | – |
| 14 | AH | – | .86 | – | – | .06 | .84 | .02 | −0.10 | – | **.87** | – | – |
| 15 | IN | – | – | .40 | – | .25 | .15 | .33 | .09 | – | – | **.63** | – |
| 16 | SD | – | – | – | **.65** | – | – | – | .64 | – | – | – | .60 |
| 17 | EX | .56 | – | – | – | **.59** | −0.09 | .09 | .05 | .46 | – | – | – |
| 18 | AH | – | .77 | – | – | .00 | .76 | .04 | .16 | – | **.79** | – | – |
| 19 | IN | – | – | **.60** | – | – | – | .53 | – | – | – | .32 | – |
| 20 | SD | – | – | – | .70 | −0.16 | .15 | .11 | .68 | – | – | – | **.83** |

The largest loading value for each item is written in bold.

**Table 9.** Inter-factor correlation matrix in the second subsample obtained with MR-CFA, ESEM-IC and R-CFA.

| FACTORS | MR-CFA | | | | ESEM-ICB | | | | R-CFA | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EX | AH | IN | DS | EX | AH | IN | DS | EX | AH | IN | DS |
| EX | 1 | | | | 1 | | | | 1 | | | |
| AH | .40 | 1 | | | .43 | 1 | | | .37 | 1 | | |
| IN | .57 | .31 | 1 | | .45 | .02 | 1 | | .69 | .42 | 1 | |
| SD | – | – | – | 1 | – | – | – | 1 | – | – | – | 1 |

had been analyzed could be expected in the population. The sample analyzed was the second subsample.

Although the analysis was carried out with the same sample as the one analyzed with MR-CFA, the goodness-of-fit indices suggest a worse fit: Chi-squared/$gl = 2.97$, $RMSEA = 0.076$, $CFI = 0.940$ and $GFI = .950$. The loading matrix obtained after R-CFA is shown in Table 3, and the inter-factor correlation matrix is shown in Table 4. In general, the loading values obtained by the different methods do not substantially differ: the most important differences are that, quite systematically, R-CFA estimated larger loading values. This tendency to obtain larger estimates could be a method effect. The same effect can be observed when the inter-factor correlation matrices in Table 4 are compared.

### 5.5. Exploratory Structural Equation Modeling, Independent-Clusters-Basis (ESEM-ICB) of the second subsample

To compare the outcome obtained with MR-CFA against the outcome obtained with a ESEM approach, we fitted an ESEM-ICB. Again, the sample analyzed was the second subsample. Items 1, 2, 5, 6, 12, 16, and 19 were used as indicators. Please note that only one indicator was defined for factor IN because in the previous EFA we could only identify an optimal indicator (item 19) for this factor. The inter-factor correlation matrix was also constrained so that the fourth factor (*SD*) was orthogonal to the other three factors, and a residual correlation was also fixed to the value estimated in the previous EFA. Four factors were extracted using the Unweighted Least

Square method. The loading matrix obtained after ESEM-ICB is shown in Table 8, and the inter-factor correlation matrix is shown in Table 9.

The goodness-of-fit indices suggest that the fit was appropriate: *Chi-squared/gl* = 1.16, *RMSEA* = 0.022, *CFI* = 0.996, and *GFI* = .994. As could be expected, this approach achieved the best goodness-of-fit, and agrees with the conclusion obtained by MR-CFA that the model can be expected at the population level.

## 6. Implementation

We implemented mild-restricted confirmatory factor analysis in R (R script "MildRestrictedCFA.r"). This script uses only native functions in R, so no packages need to be downloaded. In addition, we implemented the full item selection proposal in our factor analysis program, which can be downloaded for free from the site https://www.psicologia.urv.cat/ca/utilitats/mild-restricted-confirmatory-factor-analysis/.

## 7. Discussion

We propose a new methodological approach to compute R-CFA that aims to avoid biases in the estimated parameters due to the misspecifications in the proposed factor model. The key idea is that free parameters need to be estimated using unrestricted techniques before the restrictions are applied. In the simulation study, we compared the outcomes of our approach to those of R-CFA. MR-CFA achieved less biased estimates of the model parameters that were not restricted. In addition, the goodness-of-fit indices suggested that the model fit was better. In the analysis of a real dataset, we proposed a population factor model with several restrictions in order to show that the method can be used to assess complex population models. Again, RM-CFA gave a better model fit than the restricted-factor analysis.

While restricted CFA is nowadays considered as a less suitable analysis technique (Ferrando and Lorenzo-Seva 2000; Steenkamp and Maydeu-Olivares 2023), in the present manuscript we focused on the fact that it can be useful as long as the free parameters are correctly estimated. When CFA is computed with the methods currently available, misspecifications of zero loadings lead to distorted factors with over-estimated factor parameters and subsequent distorted structural relations (Asparouhov and Muthén 2009). The solutions proposed so far in the literature are in the line of computing less restrictive factor models, such as applying an oblique rotation to a partially specified target (Browne 1972) or the Independent-Clusters-Basis (ICB) model proposed by McDonald (1999, p 181). Our proposal is a two-step approach: in the first step, the free parameters are estimated; and in the second step, the constraints are applied. In this way, our approach aims to be able to assess CFA models without the risk of obtaining distorted parameter estimates. The final objective is to make CFA a suitable factor technique.

We tested MR-CFA based on a simple simulation study. As the differences between MR-FCA and CFA have already been pointed out, we did not carry out an extensive simulation study. However, it could be interesting to study further the performance of MR-CFA related to different situations (for example, different sample sizes, different numbers of variables, different numbers of factors, different levels of values in the secondary loadings, different levels of correlation among the factors, etc.). This kind of intensive simulation study might help to better understand the potential limitations of MR-CFA when real datasets are analyzed.

We proposed and tested MR-CFA for a single sample. Further research should be carried out to extend our approach to multiple samples. In this sense, the MR-CFA could be generalized to analyze covariance matrices (not just correlation matrices), so that factor invariance could also be explored using MR-CFA.

A limitation that the new approach is that it is computationally intensive, and it takes more time than other approaches to be computed. We must add that this should not be a serious drawback, because computer technology advances so fast that the computing time that could be considered as excessive today, will be affordable in a short time.

Further research should also be done to assess whether response styles can negatively affect the assessment of confirmatory factor models. Voss (2024) observed that careless responses deteriorate model fit under most circumstances. It could be researched whether these kinds of responses also have a negative effect when applied to MR-CFA.

### 7.1. Practical recommendations for applied researchers

The most important recommendation for applied researchers is that CFA should only be applied when the aim is to assess a hypothesized model. CFA should not be used as an exploratory technique. When a factor model is available, MR-CFA can be used to assess whether it can be expected to hold at population level. If a factor model is not available or the model available is already suspected to be erroneous, then EFA is the most recommended approach instead of CFA.

A researcher might decide to assess a factor model that has already been studied by previous research using CFA. Unfortunately, the outcome of this analysis can suggest that the model is suspicious of not being true at the population level. If the RM-CFA concludes that the population model proposed by the researcher does not hold at the population level, we would not advise that 'ad hoc' modifications be introduced into the specified factor model. Instead, we would recommend computing a full exploratory factor analysis to study the factor model suggested by the data, and also obtaining a new sample to compute a new MR-CFA that helps to assess whether the factor model suggested by the data in the first sample holds at the population level.

In our manuscript, we have always computed MR-CFA related to factor models with perfect indicator items (i.e. each item is related to a single factor, and the secondary factor loading values are set to zero in the hypothesized factor model). We have analyzed this kind of factor model because it is the most demanding and difficult model to manage from a computational point of view. Usually, the assessment of this kind of highly restricted factor model leads to a distorted estimation of the factor parameters. In our manuscript, we aimed to show that our MR-CFA can deal with such a complex model, and that R-CFA fails to do so. However, in applied research there is no need to compute MR-CFA based on such restricted factor models. If previous exploratory factor analyses (computed with different samples) suggest that some secondary loading values are actually meaningful (and they have a substantive interpretation), then a more relaxed model could be proposed in which some items are not expected to be perfect indicators of a particular factor. In this sense, MR-CFA can also be used to assess factor models that are close to the Independent-Clusters-Basis (ICB) model proposed by McDonald (1999, p 181).

A final caution regarding residuals must be considered, as they are critical for detecting local misfit (as opposed to global misfit). Theoretical expectations suggest that R-CFA is likely to produce smaller residuals, even though MR-CFA provides less biased estimates of substantive loadings. This outcome arises because R-CFA seeks an optimal solution, whereas MR-CFA constrains the substantive loadings to the values estimated during the first step using EFA. Consequently, MR-CFA will tend to exhibit smaller residuals for specific clusters of items (those loading on the same factor) but higher residuals on average. If MR-CFA yields large residuals, this could indicate that high cross-loading values have been inadvertently overlooked. Researchers are therefore responsible for proposing factor models, informed by substantive theoretical frameworks or prior analyses, that account not only for the most salient loading values per factor but also for significant cross-loadings.

## Acknowledgments

This paper is dedicated to John Bates, who has been my linguistic assessor since I began publishing in 1992. Now that he has retired, writing papers will never be as much fun.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This project has been made possible by the support of the Ministerio de Ciencia e Innovación, the Agencia Estatal de Investigación (MICIN/AEI/10.13039/501100011033) (PID2023-148374NB-I00).

## Data availability statement

Research data for the illustrative example can be obtained from the open data repository CORA: https://dataverse.csuc.cat/dataset.xhtml?persistentId = doi:10.34810/data1075.

## References

Asparouhov, T., and B. Muthén. 2009. Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal* 16 (3):397–438. doi:10.1080/10705510903008204.

Bollen, K. A. 1989. *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., and R. A. Stine. 1992. Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research* 21 (2):205–29. doi:10.1177/0049124192021002004.

Browne, M. W. 1972. Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology* 25 (2):207–12. doi:10.1111/j.2044-8317.1972.tb00492.x.

Cohen, J. 1998. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ferrando, P. J. 2021. Seven decades of factor analysis: From Yela to the present day. *Psicothema* 33 (3):378–85. doi:10.7334/psicothema2021.24.

Ferrando, P. J., and U. Lorenzo-Seva. 2000. Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica* 21 (2):301–23.

Ferrando, P. J., and U. Lorenzo-Seva. 2021. The appropriateness of sum scores as estimates of factor scores in the multiple factor analysis of ordered-categorical responses. *Educational and Psychological Measurement* 81 (2):205–28. doi:10.1177/0013164420938108.

Ferrando, P. J., A. Hernandez-Dorado, and U. Lorenzo-Seva. 2022. Detecting correlated residuals in exploratory factor analysis: New proposals and a comparison of procedures. *Structural Equation Modeling: A Multidisciplinary Journal* 29 (4):630–8. doi:10.1080/10705511.2021.2004543.

Ferrando, P. J., A. Hernández-Dorado, and U. Lorenzo-Seva. 2024. A simple two-step procedure for fitting fully unrestricted exploratory factor analytic solutions with correlated residuals. *Structural Equation Modeling: A Multidisciplinary Journal* 31 (3):420–8. doi:10.1080/10705511.2023.2267181.

Hu, L. T., and P. M. Bentler. 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. doi:10.1080/10705519909540118.

Jennrich, R. I., and P. F. Sampson. 1966. Rotation for simple loadings. *Psychometrika* 31 (3):313–23. doi:10.1007/bf02289465.

Jöreskog, K. G. 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34 (2):183–202. doi:10.1002/j.2333-8504.1967.tb00991.x.

Li, C. H. 2016. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods* 48 (3):936–49. doi:10.3758/s13428-015-0619-7.

Lorenzo-Seva, U. 2022. SOLOMON: A method for splitting a sample into equivalent subsamples in factor analysis. *Behavior Research Methods* 54 (6):2665–77. doi:10.3758/s13428-021-01750-y.

Lorenzo-Seva, U., and P. J. Ferrando. 2019. Robust Promin: A method for diagonally weighted factor rotation. *Liberabit: Revista Peruana de Psicología* 25 (1):99–106. doi:10.24265/liberabit.2019.v25n1.08.

Lorenzo-Seva, U., and P. J. Ferrando. 2020. Unrestricted factor analysis of multidimensional test items based on an objectively refined target matrix. *Behavior Research Methods* 52 (1):116–30. doi:10.3758/s13428-019-01209-1.

Lorenzo-Seva, U., and P. J. Ferrando. 2021. Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution. *Structural Equation Modeling: A Multidisciplinary Journal* 28 (1):138–47. doi:10.1080/10705511.2020.1735393.

Lorenzo-Seva, U., and P. J. Ferrando. 2023. A simulation-based scaled test statistic for assessing model-data fit in least-squares unrestricted factor-analysis solutions. *Methodology* 19 (2):96–115. doi:10.5964/meth.9839.

Lorenzo-Seva, U., A. Vigil-Colet, and P. J. Ferrando. 2024. Spanish sample responses to the Statistical Anxiety Scale psychological test. CORA. Repositori de Dades de Recerca.

Lorenzo-Seva, U., A. Vigil-Colet, and P. J. Ferrando. 2022. Development of a revised version of the statistical anxiety scale. *Psicothema* 34 (4):562–70. doi:10.7334/psicothema2022.231.

Maydeu-Olivares, A., C. Ximénez, and J. Revuelta. 2024. Percentage of variance accounted for in structural equation models: The rediscovery of the goodness of fit index. *Psychological Methods* doi:10.1037/met0000680.

McDonald, R. P. 1999. *Test theory: A unified treatment*. Hillsdale, MI: LEA.

Nájera, P., F. J. Abad, and M. A. Sorrel. 2023. Is exploratory factor analysis always to be preferred? A systematic comparison of factor analytic techniques throughout the confirmatory–exploratory continuum. *Psychological Methods* doi:10.1037/met0000579.

Stankov, L., R. Roberts, and G. Spilsbury. 1994. Attention and speed of test-taking in intelligence and aging. *Personality and Individual Differences* 17 (2):273–84. doi:10.1016/0191-8869(94)90031-0.

Steenkamp, J. B. E., and A. Maydeu-Olivares. 2023. Unrestricted factor analysis: A powerful alternative to confirmatory factor analysis. *Journal of the Academy of Marketing Science* 51 (1):86–113. doi:10.1007/s11747-022-00888-1.

Steiger, J. H. 2007. Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences* 42 (5):893–8. doi:10.1016/j.paid.2006.09.017.

Tabachnick, B. G., and L. S. Fidell. 2007. *Using multivariate statistics*. 5th ed. New York: Allyn and Bacon.

Timmerman, M. E., and U. Lorenzo-Seva. 2011. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological methods* 16 (2):209–20. doi:10.1037/a0023353.

Voss, N. M. 2024. The effects of careless responding on the fit of confirmatory factor analysis and item response theory models. *Behavior Research Methods* 56 (2):577–99. doi:10.3758/s13428-023-02074-9.