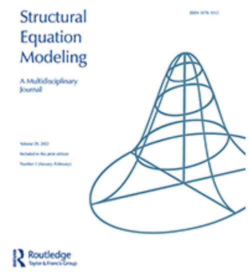

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Detecting Correlated Residuals in Exploratory Factor Analysis: New Proposals and a Comparison of Procedures

Pere J. Ferrando, Ana Hernandez-Dorado & Urbano Lorenzo-Seva

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2021.2004543

Routledge
Taylor & Francis Group

TEACHER'S CORNER

OPEN ACCESS Check for updates

Detecting Correlated Residuals in Exploratory Factor Analysis: New Proposals and a Comparison of Procedures

Pere J. Ferrando, Ana Hernandez-Dorado, and Urbano Lorenzo-Seva

Universitat Rovira I Virgili

ABSTRACT

In the classical exploratory factor analysis (EFA) model, residuals are constrained to be uncorrelated. However, since the 1960s, extensions of the classical model that allow correlated residuals to be modeled exist. Furthermore, in many EFA applications (especially those intended for item analysis) it is highly relevant to decide whether an extended solution is more appropriate than the simpler classical solution. This decision, in turn, requires effective and powerful methods for detecting correlated residuals (doublets) when they are really present to be available. This paper discusses two existing detection approaches in the EFA context, and proposes a third, new procedure. Reference values, based on the concept of parallel analysis, are proposed for deciding the relevance of the flagged doublets in all the considered procedures. The functioning of the three procedures is assessed by using simulation, and illustrated with an illustrative example. The proposal, finally, has been implemented in a well-known noncommercial EFA program, and an implementation in R is being developed.

KEYWORDS

Exploratory factor analysis; correlated residuals; doublets; expected parameter change; image theory; parallel analysis

A distinctive feature of the restricted (confirmatory) factor analysis model (CFA) is that it allows correlated residuals to be specified. In contrast, in the unrestricted (exploratory) FA (EFA) model, the residual matrix is assumed to be diagonal, and so, all the residual correlations are constrained to be zero. This distinction, however, only holds for the "classical" EFA model (e.g., Mulaik, 2010; Sörbom, 1975), as, since the 1960s, more flexible EFA solutions that allow correlated residuals to be modeled have been proposed (Butler, 1968; McDonald, 1969; Mulaik, 2010; Yates, 1987). At present, correlated residuals within an EFA solution can be modeled via ESEM (Asparouhov & Muthén, 2009; Van Kesteren & Kievit, 2020), or R packages, such as the function "esem" in the "psych" package (Revelle, 2021) or combining the packages "psych" and "lavaan" (Rosseel, 2012).

There is a vast literature and a heated debate on the convenience of allowing correlated residuals to be modeled in FA solutions (e.g., Asparouhov et al., 2015; MacCallum et al., 1992). Given the aims of this article (which are stated below), an exhaustive review of these issues is beyond its scope. However, an initial, brief discussion as well as a statement of our perspective is in order.

In an EFA based on sample data, detected correlated residuals might occur for at least two reasons: First is "true" population residuals that are perfectly justifiable (e.g., because of repeated presentation of the same items, wording similarities in the item stems, context effects … etc.). Second is sampling fluctuation around "true" zero residuals. In the first case, constraining these residuals to be zero are specification errors that, in principle, are expected to give rise to two consequences: (a) bad model-data fit, and (b) biased parameter estimates (e.g., Montoya & Edwards, 2021; Mulaik, 2010; Yates, 1987). As for point (a), however, in many applications, reasonably acceptable fits might still be obtained, but at the cost of grossly biased parameter estimates or additional factors that do not reflect substantive content (e.g., Costner & Schoenberg, 1973).

In the case of spurious correlated residuals that are due to sampling error, freeing them "blindly" with the sole purpose of improving model-data fit, is a clear case of capitalization on chance. Model-data fit in this particular sample will, no doubt, be improved, but results would not reflect "true" model-data fit in the population. Furthermore, biased estimates would be also expected to occur. However, the literature suggests that the biases in this case are not as large as those expected when "true" residuals are constrained to be zero (Reddy, 1992). We believe that blind post-hoc modification with the sole purpose of improving fit is unacceptable practice (e.g., Browne, 2001). So, the tools we shall propose in this paper are expected to be used in scenarios in which (a) samples are large enough to reasonably prevent substantial spurious correlated residuals to appear, and (b) the detected correlated residuals have a clear and defensible justification. Furthermore, we believe that cross-validation must become a routine practice for checking the consistency and stability of the diagnostic results. Having said that, we believe that obtaining efficient and powerful EFA-based diagnostic procedures for detecting correlated residuals is a relevant aim and has a clear interest.

The present article considers and compares three approaches (and indices derived from them), aimed at detecting the presence and magnitude of correlated residuals in datasets that are to be fitted by a classical EFA solution. Of these procedures, the first is the most common, and is

CONTACT Pere J. Ferrando perejoan.ferrando@urv.cat Facultad de Psicología, Universidad "Rovira I Virgili", Carretera Valls S/n, Tarragona 43007, Spain.

implemented in most EFA packages. The second, has been proposed in the literature, but is less known. The third approach, finally is, we believe, a new contribution. Furthermore, a new general procedure for obtaining reference or cutoff points for all the indices is proposed. So, this article not only reviews and compares existing procedures, but aims also to make new methodological contributions. Finally, an instrumental contribution of this proposal is that all the considered indices and reference values have been implement in a noncommercial widely used EFA program.

A variety of dimensional-reduction procedures that allow residual correlations to be detected under fixed conditions (i.e. when conditional independence would be expected), and that go beyond the specific domain of FA have been proposed in the literature (e.g., Glymour et al., 2019). Within the FA modeling, correlated-residuals-diagnostic procedures have been proposed in the CFA context, and, of these, perhaps, the most well-known are those based on Bayesian analysis (BSEM; Asparouhov et al., 2015; Zhang et al., 2021). None of these approaches will be discussed here; however, as we intend to focus only in EFA solutions.

Overall, the present proposals are thought to be particularly useful in calibration-type psychometric applications in which item scores are factor-analyzed. On the one hand, in most cases item banks are large and item scores have complex structures (Cattell, 1952; Marsh et al., 2014), which makes unrestricted FA quite an appropriate model (P.J. Ferrando & Lorenzo-Seva, 2000). On the other hand, as mentioned above, correlated residuals are very common with this type of variables. So common, in fact, that in the context of item factor analysis, they traditionally receive the specific name of "doublets" (Mulaik, 2010; Thurstone, 1947) a name that we shall also use here. However, we note that the present proposal might be used in different types of applications, such as longitudinal studies (e.g., Little, 2013).

Rationale, basic results, and estimation procedures

As a basis for our proposal, we shall first consider an extended, unrestricted multiple FA solution in which the initial unrotated pattern loading matrix is in canonical form:

$$\mathrm{R} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}\mathbf{R_{uu}}\boldsymbol{\Psi} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \mathbf{C_{uu}}. \tag{1}$$

where $\mathbf{R}$ is the $m \times m$ inter-item correlation matrix, $\boldsymbol{\Lambda}$ is the $m \times r$ canonical unrotated pattern (e.g., Harman, 1976), $\boldsymbol{\Psi}$ is the $m \times m$ diagonal matrix containing the item residual standard deviations, and $\mathbf{R_{uu}}$ is the $m \times m$ residual correlation matrix. So, $\mathbf{C_{uu}} = \boldsymbol{\Psi}\mathbf{R_{uu}}\boldsymbol{\Psi}$ is the residual covariance matrix.

If the residuals are all uncorrelated, $\mathbf{R_{uu}}$ becomes an identity matrix, and model (1) reduces to the classical EFA model

$$\mathrm{R} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}^2. \tag{2}$$

Solutions (1) and (2) are direct, unrotated solutions, and are expected to be further transformed (possibly obliquely) at the rotation stage. Transformations of the common part of the solution, however, do not affect the detection of correlated residuals. For this reason, we consider the most usual, and possibly simpler, direct orthogonal solution (the canonical pattern) as a basis for our proposal. We also assume $\mathbf{R}$ to be positive definite (see Lorenzo-Seva & Ferrando, 2021). However, apart from that, the basic modeling is quite comprehensive, and can be considered for (a) binary scores, (b) graded scores treated as ordered categorical variables, and (c) graded or more continuous scores treated as continuous variables. In case (a) the elements of $\mathbf{R}$ are tetrachoric correlations. In case (b) they are polychoric correlations, and in case (c) they are product-moment (Pearson) correlations.

From the modeling bases so far described, detection of correlated residuals can be viewed as an assessment of the extent to which the simpler solution (2) is appropriate with respect to the extended solution (1). In more detail, what is to be assessed is (a) the magnitude of the non-diagonal elements of $\mathbf{R_{uu}}$, and (b) the extent to which constraining these elements to be zero, biases the structural estimates of the common-factor parameters in (2), (i.e. loadings and residual variances).

Three approaches for detecting the presence of correlated residuals

Suppose that the extended solution (1) holds for a given set of measurements but it is the classic solution (2) that is instead fitted to a sample inter-item correlation matrix. The elements of the fitted residual matrix

$$\mathrm{C_{res}} = \hat{\mathrm{R}} - \hat{\Lambda}\hat{\Lambda}', \tag{3}$$

represent the differences between the observed sample correlations and the model-fitted, or model-expected correlations. These elements are usually known as *correlation residuals* (Bollen, 1989) but this name is potentially misleading, because they are in fact covariances (see Equation (1)).

At first sight, the $\mathbf{C_{res}}$ matrix can be considered to be an estimate of the "true" $\mathbf{C_{uu}}$ matrix in (1). So, the largest (in absolute value) elements of $\mathbf{C_{res}}$ would indicate which the most salient doublets are. This approach is the most common in practice, and is the first that we shall consider here.

Evidence suggests that in some cases (particularly in simple and clear solutions) the standard approach based on (3) works well, but in others might be highly misleading (Blalock, 1971, part IV; Costner & Schoenberg, 1973; Sörbom, 1989; Little, 2013). If substantial correlated residuals exist but are forced to be zero, the un-modeled or omitted correlations or covariances will tend to be "re-assigned" in parameter estimation so as to keep the omitted parameter as close to zero as possible, which means that the estimated $\boldsymbol{\Lambda}$ loadings as well as the remaining residual estimates are expected to be biased to a greater or lesser extent. So, we might as well end with a substantially biased estimated solution (most Heywood cases are due to doublets; see McDonald, 1985) while, at the same time, the "true" culprit residuals remain unsuspectedly low in the fitted matrix (3).

The second approach we shall consider can be derived from Guttman's (1953) image theory. The negatives of the off-diagonal elements of the anti-image correlation matrix contain the partial correlations between the corresponding pairs of variables after conditioning on the remaining variables. These elements are available in several standard statistical packages as indices for assessing "sampling adequacy"–the extent to which the data fulfil preliminary assumptions of the EFA model–(Kaiser, 1974). Here, however, we propose to use them for different purposes (Mulaik, 2010, p. 233). Define:

$$\begin{aligned} S^2 &= [diag(R^{-1})]^{-1} \\ Q &= SR^{-1}S. \\ P &= 2I - Q \end{aligned} \tag{4}$$

Then **Q** is the anti-image correlation matrix, and **P** the partial correlation matrix (with unit values in the main diagonal). Now, if model (1) is correct, then **P** is an estimate of the "true" $\mathbf{R_{uu}}$ correlation matrix in (1), and will approach more and more $\mathbf{R_{uu}}$ as the ratio variables to factors increases without bound (Kaiser, 1963; McDonald, 1985; Mulaik, 2010).

As a diagnostic tool for detecting doublets, **P** has two obvious advantages. First, it is obtained "a priori" with no need to fit first any EFA model. Second, its elements are readily interpretable as correlations. As for the shortcomings, even when the elements of **P** are obtained "a priori," they are only correct estimates of $\mathbf{R_{uu}}$ for the solution with the correct number of common factors. Second, they only become really interchangeable to their corresponding parameter values as the number of indicators per factor is very large. In solutions with few indicators per factor, the off-diagonal elements of **P** are expected to overestimate their corresponding true values (Kaiser, 1963; McDonald, 1985, p. 72), which would lead to detect more doublets than there really are. Our simulations below clearly support this expectation.

The third diagnostic tool that we propose appears to be a new contribution, and is derived from the concept of "Expected Parameter Change" (EPC; Saris et al., 1987, 2009). We shall first describe its general rationale, and then describe the specific estimation procedure we propose.

The EPCs in our proposal are obtained sequentially as follows: for each of the *(m×(m–1))/2* possible doublets, the constraint: ρ_{ujuk}=0 (which involves variables *j* and *k* that form the particular possible doublet under study) is relaxed, and this residual correlation is freely estimated by using the estimation procedure described below. Furthermore, the vectors of loadings corresponding to variables *j* and *k* are estimated (a) under the standard ρ_{ujuk}=0 constraint, and (b) when the residual correlation is freely estimated. Now, if we denote the reproduced communality of variable *j* as $\hat{\mathbf{h}}^2_j = \sum_{q=1}^{r} \hat{\lambda}^2_{jq}$, the two types of EPC we propose for each possible pair of variables are as follows :

EPC1: **E**xpected **RE**sidual correlation direct **C**hange index (*EREC* index)

$$EPC1_{j,k} = |\hat{\rho}_{ujuk} - 0|. \tag{5}$$

EPC2: **E**xpected commu**N**al**I**ty **D**ir**E**ct change Index (*ENIDE* index)

$$EPC2_{j,k} = \frac{|(\hat{h}_j^{2(0)} - \hat{h}_j^{2(1)})| + |(\hat{h}_k^{2(0)} - \hat{h}_k^{2(1)})|}{2}. \tag{6}$$

The interpretation of both indices is rather simple. *EREC* quantifies the amount of misspecification in the residual correlation itself, which is induced by setting ρ_{ujuk}=0, and its magnitude is interpreted as a correlation coefficient. As for *ENIDE*, it quantifies in a single index the extent to which this misspecification "propagates" and produces biased loading estimates, and is interpreted as an average proportion of change (in the estimated common variance of both variables).

We turn now to the proposed estimation procedure for the tool so far summarized. It is based on previous proposals by Wright (1968) and Yates (1987), who called them *residual omission* and *sectioning*, respectively, and is intended to minimize the biases of both the residual estimates themselves and the rest of the structural parameter estimates in $\mathbf{\Lambda}$ and $\mathbf{\Psi}$. Essentially, it is a three-stage unweighted least squares (ULS) procedure. In the first stage, and for each possible doublet (*j*, *k*) the implied variables are omitted from the dataset, and model (2) is fitted to the remaining (*m*–2) variables (that we refer as the core set) by using ULS estimation. If the pair (*j, k*) is, actually, a nontrivial doublet this first-stage fitting is expected to provide less biased estimates of the elements of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ corresponding to the core variables. We note also that, once a possible doublet has been omitted, at least three variables must remain if a factor solution is to be fitted to the core set. So, even in the simplest unidimensional case, the estimation procedure is not feasible with less than five variables.

At the second stage, the estimates of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ for the two variables not included in the first-stage core set, are obtained by using extension analysis (e.g., Nagy et al., 2017) separately for each variable. Thus, for the *j* variable omitted in step 1, the pending estimates are $\boldsymbol{\lambda}_j$ and $\varphi_{j,,}$, and are obtained as follows: Let $\mathbf{r}_{j,core}$ be the column vector containing the correlations between variable *j* and the variables in the core set. The pending estimates are obtained as

$$\hat{\lambda}_j = (\hat{\mathbf{\Lambda}}_{core}{}'\hat{\mathbf{\Lambda}}_{core})^{-1}\hat{\mathbf{\Lambda}}_{core}{}'\mathbf{r}_{j,core} \; ; \; \hat{\phi}_j = \sqrt{1 - \hat{\lambda}_j{}'\hat{\lambda}_j} \tag{7}$$

At the end of stage 2 estimates of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ for the full set of variables have been obtained. At the third stage, finally, the estimate of the residual correlation specified to be free is obtained by

$$\hat{\mathbf{R}}_{\mathbf{uu}} = \hat{\mathbf{\Psi}}^{-1}(\mathbf{R} - \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}')\hat{\mathbf{\Psi}}^{-1}. \tag{8}$$

And the (*j,k*) element of $\hat{R}_{uu}$ in (8) is the estimate we want to obtain. As stated above, the process so far described is carried out for each possible doublet until all the non-duplicated non-diagonal elements of $\hat{R}_{uu}$ have been estimated. Overall, $\hat{R}_{uu}$ is intended to be an estimate of $\mathbf{R_{uu}}$ which is obtained from estimates of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ that have been corrected for potential biases due to the "propagating" effects of the non-modeled doublets.

We shall name this estimation procedure **M**inimum expected bias **O**f **R**esidual and loadin**G** v**A**lues i**N** s**A**mple estimates method (MORGANA method).

The choice of the ULS criterion in MORGANA can be justified on several grounds. ULS is easily implemented and computationally robust (Forero et al., 2009; Jöreskog, 2003; Lee et al., 2012; Mislevy, 1986; Zhang & Browne, 2006), and this is particularly relevant here given that the proposal is expected to be used in datasets with a large number of variables and complex structures. Furthermore, in many cases, it will be used with solutions based on tetrachoric and polychoric **R**'s, in which ULS has proved to be a defensible option (Forero et al., 2009; Knol & Berger, 1991; Lee et al., 2012; Zhang & Browne, 2006).

Reference values and upper bounds

The outcome of any of the procedures considered here would consist of *(m×(m–1))/2* estimated indices, whose magnitude is thought to indicate the plausibility that the corresponding residual correlation is a true doublet. In order to make correct decisions, however, reference values for deciding whether the obtained index flags a true doublet or is a false alarm are clearly needed. The general approach we propose for obtaining these values is based on the concept of Parallel Analysis (PA; see, e.g., Timmerman & Lorenzo-Seva, 2011) and is discussed below in relation to the different procedures.

EREC, fitted residuals, and partial correlations

Indices in this group are estimates of the elements of the residual covariance matrix $\mathbf{C}_{\mathbf{uu}}$ (fitted residuals), or of the residual correlation matrix $\mathbf{R}_{\mathbf{uu}}$ in (1). So the basis procedure is the same for all of them. Let **X** ($N \times m$) be the data matrix that is factor-analyzed. The first step is to construct the matrix **Xa** ($N \times m$) in which each column has been obtained from the corresponding column in **X** but with the elements re-ordered at random. So, the distribution of the columns of **X** and **Xa** is the same, but the correlations between the columns of **Xa** are all zero (at least in the population). Let now be **Ca** the covariance matrix, **Ra** the correlation matrix, and **Pa** the partial correlation matrix obtained all of them from **Xa**. The elements of **Ca** and **Ra** can be viewed as the residual covariances and correlations respectively that would arise merely by sampling error, and so, are appropriate references against which the fitted residuals and the EPC1 values (respectively) obtained from **X** can be compared. The elements of **Pa** can be viewed as the partial correlations that would arise solely because of sampling error, and so can be taken as appropriate references against which the corresponding values in **P** can be compared.

The process of creating **Xa** from **X** is repeated 500 times and this provides a distribution of values for each element of **Ra** and **Pa**. Typically, in PA the mean and the 95$^{\text{th}}$ centile are considered as suitable thresholds. In the simulation study that follows, we shall test both of them. In order to do it, the mean of the distribution (as well as the 95$^{\text{th}}$ centile) is computed from the random distribution of each element of **Ra** and **Pa**. With the aim of having a single value as threshold, the grand mean is computed and used as criterion (i.e., the mean of the means, and the mean of the 95$^{\text{th}}$ centiles). We can advance that both thresholds lead to very similar outcomes, with that based on the means being the most accurate.

ENIDE

ENIDE values are estimates of average communality changes, and the procedure for obtaining reference values is not as direct as above. Let **X** ($N \times m$) be as before, and $\hat{\mathbf{\Lambda}}$ ($N \times r$) the estimated factor pattern matrix obtained from factorizing **X**. We first generate a dataset in which $\hat{\mathbf{\Lambda}}$ is the "true" population pattern and the residuals are uncorrelated. To do so, we (a) compute the reproduced correlation matrix $\mathbf{R}^* = \hat{\mathbf{\Lambda}}\,\hat{\mathbf{\Lambda}}'$ and obtain **L**, the Cholesky decomposition of $\mathbf{R}^*$. And (b) generate a matrix $\mathbf{Z}(N \times r)$ of uncorrelated standard normal scores (in the population) and compute $\mathbf{Y} = \mathbf{ZL'}$. Then Y is a matrix of factor scores derived from a solution in which $\hat{\mathbf{\Lambda}}$ is true in the population and the residuals are uncorrelated.

At the next stage, we obtain the correlation matrix of the **Y** scores, factorize this matrix in r factors, and obtain an estimated pattern denoted by $\hat{\mathbf{\Lambda}}_k$. When ENIDE is computed from $\hat{\mathbf{\Lambda}}_k$ we are obtaining the communality changes that can be expected solely by chance. As in the proposal above, the process is repeated 500 times and this provides a distribution of values for each ENIDE that would be obtained if no doublets existed in the solution and estimates of parameter change were solely due to sampling error. As in the previous case, we consider as possible thresholds the mean and the 95$^{\text{th}}$ centile of the random distributions.

An illustrative example

The small example described in this section is expected to be useful for illustrating the rationale of the proposal, as well as the functioning of the methods that are to be compared. Suppose that a researcher wants to analyze the scores on a 6-item set expected to measure a single dimension. However, the wording of items 1 and 2 is similar, leading to a substantial doublet. We produced an artificial dataset with the item scores of 500 individuals to these 6 items. As we generated the data, we know that the correlation residual between item 1 and 2 is .40 at the population. Table 1 shows the sample correlation matrix that is positive definite and well suited for FA (Lorenzo-Seva & Ferrando, 2020). We fitted a single-factor ULS solution, and obtained the results in Table 1.

The estimated loadings show substantial biases with respect to the true loadings. The loadings in the two items that form the doublet are clearly inflated, whereas the remaining loadings are strongly deflated. Note that the loading corresponding to the first item of the doublet even approaches a Heywood case. The researcher, who is not aware of the existence of the doublet, would likely think that item 1 is an excellent indicator of the common factor: almost a marker in fact.

In order to assess now the possible existence of doublets, we first inspect the fitted-residual-matrix $\mathbf{C}_{\mathbf{res}}$ in Equation (3). It is in Table 2:

In this example $\mathbf{C}_{\mathbf{res}}$ would do a poor job in identifying the "true" doublet involving items 1 and 2, which remains undetected. In contrast, the cutoff value based on PA

Table 1. Sample correlation matrix and loading matrices related to the illustrative example.

Items	Sample correlation matrix						ULS-EFA	True
	1	2	3	4	5	6	Loadings	Loadings
1	1						0.902	0.60
2	.688	1					0.684	0.60
3	.275	.192	1				0.366	0.60
4	.264	.128	.224	1			0.308	0.60
5	.278	.204	.251	.171	1		0.357	0.60
6	.256	.204	.157	.119	.128	1	0.311	0.60

Table 2. Fitted-residual matrix. Illustrative example.

Items	Fitted-residual matrix					
	1	2	3	4	5	6
1	-					
2	.071	-				
3	−.056	−.059	-			
4	−.014	−.082	.111*	-		
5	−.045	−.040	−119*	.060	-	
6	−.026	−.008	.043	.023	.017	-

would signal the residual covariances 3–4 and 3–5 (marked with an *) as possible doublets. The misspecification error in this case has clearly propagated, and the omitted covariance has been "re-assigned" to the loading estimate of the first item, and also to other residual covariances, whose estimated values are far larger than that corresponding to the original doublet.

Let us try now the second procedure. The anti-image-based partial correlation matrix is in (Table 3):

Clearly**, P** does here a far better job than $\mathbf{C_{res}}$, and its largest element correspond to the "true" population doublet. Note also that this value is far larger than any of the remaining estimates. However, the criterion based on PA would also flag residual 1–4 as a potential doublet. So, possibly, **P** shows here the problem derived from its asymptotic derivation, which was discussed above: In this solution with very few indicators of the factor, its elements tend to overestimate their corresponding true values, which would lead us to think that there is more than one substantial doublet in this solution.

Table 3. Anti-image-based partial correlation matrix. Illustrative example.

Items	Partial correlation matrix					
	1	2	3	4	5	6
1	-					
2	.652*	-				
3	.121	.008	-			
4	.192*	−.081	.143	-		
5	.131	.022	.171	.077	-	
6	.122	.043	.077	.041	.041	-

Table 4. EREC values for each possible doublet. Illustrative example.

Items	EREC					
	1	2	3	4	5	6
1	-					
2	.584*	-				
3	.269*	.178	-			
4	.072	.189	.146	-		
5	.169	.125	.169	.084	-	
6	.012	.020	.068	.033	.034	-

Table 5. ENIDE values for each possible doublet. Illustrative example.

Items	ENIDE					
	1	2	3	4	5	6
1	-					
2	.446*	-				
3	.087	.087	-			
4	.072	.056	.027	-		
5	.024	.035	.036	.017	-	
6	.181	.042	.024	.010	.018	-

We turn finally to our proposal. Tables 4 and 5 shows the EREC estimates (Equation 5) in panel (a), and the ENIDE estimates (Equation 6) in panel (b). As in the previous results, the potential doublets flagged by the PA criterion are denoted with an *.

Overall, the new proposed indices work rather well here. Both attain, bay far, the maximum value for the true doublet. EREC would still flag a non-existing second doublet, while ENIDE correctly flags the only existing doublet without arriving at any false alarm.

Additional considerations

A priori, the diagnostic procedures proposed so far are expected to work reasonably well if certain conditions are fulfilled. First, the number of common factors is correctly specified. Second, the sample is large enough to minimize large sampling fluctuations. Third, the number of substantial doublets is small relative to the size of the residual matrix $\mathbf{R_{uu}}$, (i.e. most of the correlated residuals in $\mathbf{R_{uu}}$ are near zero, and only a few are substantial). As Steiger (1990) showed, the procedures will possibly fail in a scenario in which all the residuals were correlated and the values of the residual correlations were similar among them. This scenario, however, seems quite unlikely, at least in the context of item analyses.

Within the same context of item analysis, our proposed procedures will probably work better if the few existing sizable doublets are all of the same sign (possibly positive), as this condition will prevent effects of "ensemble" biases that impact in opposite direction the estimated parameters of the FA model. In our experience, and in the item-analysis context, doublets mainly reflect shared specificities due to similarities in item content, context and/or wording (see also Bandalos, 2021). If this is so, the doublets are expected to be generally positive provided that all the items are scored in the same direction of the trait/s being measured (see Bandalos, 2021, Tables 1 and 2), which is the most usual practice. We note that the operating mechanism under this scoring will be very similar to that in repeated item presentations (including retest effects), which are expected to give rise to positive residual correlations (e.g., Little, 2013).

In order to prevent the procedures here considered to arrive at misleading detections, two basic, common-sense, initial recommendations can be made. First is to work well-designed datasets obtained from large samples. Second is to use cross-validation whenever possible. Although the PA approach described above is intended to prevent spurious detections, we agree with Steiger (1990) that, to achieve this aim, nothing substitutes replication. We would also note that both recommendations are closely linked. We would not provide here a minimal sample size criterion, because sampling stability depends on many determinants (see Lorenzo-Seva & Ferrando, 2021). Rather, our advice focus on replicability: if an FA solution (with or without doublets) is not replicated in different, well-chosen, samples, then the sample is too low.

Assuming that a correct diagnostic has been achieved, issues of model identification can be next carefully considered regarding further possible decisions made by the researcher on the basis of the detection results. Thurstone (1947) recommended to always avoid doublets, and, following this advice, the researcher might well decide to discard offending items until no doublet remains. If so, identification issues are simply those of the standard EFA model in (2) (see Hayashi & Marcoulides, 2006). On the other hand, if he/she decides to fit the more flexible model (1) (possibly in a new sample), then assessing model identification and determinacy becomes crucial, and this issue requires additional considerations above those discussed above. More specifically, the essential conditions are now two. First, as stated above, that there are only a few substantial doublets and the remaining are essentially zero. Second, that the common part of the solution is over-determined, with not too many factors and multiple indicators per factor (Hayashi & Marcoulides, 2006; Mulaik, 2010). If these conditions are not fulfilled, the resulting solution based on (1) is likely to be not unique and highly indeterminate.

In order to approach the first condition above, and also to minimize the impact of capitalization on change, for all the procedures considered in the ms., we propose to limit the maximum number of detected doublets to $g - r$, where g is Ledermann's (1937) bound solved for the number of common factors (Hayashi & Marcoulides, 2006; Mulaik, 2010). This limitation is not arbitrary, but has, in our view, a defensible rationale. Doublets can be regarded as minor factors related to a substantial amount of variance that is only shared by a pair of variables. If m is the number of observed variables, and r is the number of common factors extracted from them, then g can be interpreted as the maximum number of (major plus minor) factors that can be determined from the observed variables (or, in other words that still leaves a number of degrees of freedom greater than zero, and allows the model to be testable). Now, if we consider doublets as minor factors, then the number of doublets that can be allowed if determinacy is to be maintained is related to the specified number of common factors: So, as r approaches g, less doublets can be allowed. We would stress, however, that the restriction we impose allows a potential model of the type (1) to be identified, but does not ensure that the obtained solution is determinate. To see this point, note that the restriction only ensures that the number of degrees of freedom is above zero, but does not inform at all about whether the common factors are over-determined with a sufficient number of indicators each.

Simulation study

The simulation study summarized in this section is aimed to assess and compare the different diagnostic procedures discussed in this paper. The assessment is done in base to the sensitivity (i.e., the ability to identify in the sample data a real doublet in the population model), and specificity (i.e., the ability to identify in the sample data that a residual is not a real doublet in the population) of the diagnostic procedures. The study was based on eight independent variables, a factorial design with 3,564 conditions, and 200 replicas per condition. After each condition, the outcome among replicas was used to stablish the sensitivity and the specificity of each diagnostic method in this condition. The independent variables were:

(1) Sample size: samples were drawn to be of sizes 150, 300, and 1,000.
(2) Number of items: the number of observed variables per factor were 5 and 10.
(3) Number of factors: factor models in the population were designed so that the true number of factors were 1, 2, and 3. The number of factor extracted in the sample data corresponded always with true number of factors in the population.
(4) Inter-factor correlation: when more than one factor was present in the population factor model, orthogonal and oblique models were manipulated. In the oblique models an inter-factor correlation value of .20 was set for all the factors.
(5) The level of communality was manipulated choosing salient loading values in specific ranges. The ranges used were: .30 to .40 (low communality), .41 to .55 (medium communality), and .56 to .70 (large communality).
(6) The size of non-salient loading matrices in the population were also manipulated so that the maximum absolute values were: .05, .10, and .20.
(7) Number of doublets in the population were manipulated to be: 1, 2, 3, 4, and 5. However, the number of doublets were limited in each condition in order that the number of doublets was lower than half the number of observed variables in the factor model (for example, in the conditions in which the number of observed variables was 5, the maximum number of doublets considered was 2).
(8) Size of doublets: in the population model the size of doublets were manipulated to produce three levels: .20 to .30 (low doublets), .31 to .40 (medium doublets), and .41 to .50 (high doublets).

In all the conditions, including those in which more than one doublet were simulated in the same dataset, finally, the study included doublets in both directions. The value and the sign of each doublet were chosen in two steps: first, the value was chosen (taking in consideration the condition being simulated); second, the sign was decided at random with a chance of 50% for each sign.

The total number of samples generated in the simulation study were 712,800, and contained a total of 1,992,600 doublets. For each sample, we computed Fitted residuals, Partial correlations, EREC, and ENIDE. In order to decide which pairs should be considered as doublets in the sample, we computed PA for each diagnostic index, and registered True Positives, True Negative, False Positive, and False Negatives. After the 200 replicates of the condition at hand, we computed a ROC analysis to assess the sensitivity and the specificity of each index in this condition.

Results

In order to assess the overall performance we computed the mean and standard deviation of sensitivity and the specificity for each index among conditions. Table 6 shows these statistics.

Table 6. Mean and standard deviation of sensitivity and the specificity for each index.

Diagnostic index	Mean as threshold value of PA		C95 as threshold value of PA	
	Sensitivity	Specificity	Sensitivity	Specificity
Fitted residuals	.821 (.205)	.945 (.022)	.743 (.251)	.975 (.020)
Partial correlations	.890 (.127)	.948 (.023)	.890 (.128)	.948 (.022)
EREC	.919 (.120)	.949 (.019)	.918 (.122)	.950 (.018)
ENIDE	.692 (.147)	.944 (.017)	.661 (.157)	.951 (.013)

As can be seen in the table, mean and C95 thresholds produced similar outcomes, with the mean value providing a little more accurate diagnostics. For practical application, we would advise to use the mean as threshold.

It is interesting that, even when fitted residuals is the most frequently inspected diagnostic index by researchers, it was not the most efficient index. It must be said that when the number of items per factor was large (m/r = 10), the mean values for sensitivity and specificity were .958 and .948, respectively. However, as the number of factors increased, the sensitivity substantially decreased (mean value of .790).

Partial correlations systematically improved the performance of Fitted residuals. The conditions where its performance was most accurate were large samples, large number of items per factor, large communality, and a large size value of the doublets. In these conditions, its sensitivity and specificity were systematically over .92 and .95, respectively. In must be pointed the most factors in the factor model, the best the performance of this index.

ENIDE showed a specificity comparable to the other indices. However, its sensitivity was the worst of all of them. It must be said this index showed it best performance (Sensitivity> 0.95 and Specificity > 0.96) when N = 1,000 and a single doublet was present in the population. In addition, the errors that can affect most the estimates of a factor model are those related to the low Specificity (i.e., to fix a residual correlation to zero, when in the population the related pair is a strong doublet; Reddy, 1992).

Finally, as EREC turned out to be the most effective diagnostic index. A more acurate report of its performance is given in Table 7.

The conditions where EREC performance was most accurate were large samples, large number of items per factor, low communality, and a large size value of the doublets. In these conditions, its sensitivity and specificity were systematically over .92 and .95, respectively. It must be pointed out that the largest the communality, the worse the performance of this index is: it can be explained because when the loading value in the population model is already large, the less room there is for an overestimation of its value in the sample model.

Implementing diagnostic indices in FACTOR

The authors' experience suggests that proposals such as the present one are only used in practical applications if they are implemented in user-friendly and easily available software. In this respect, the procedure proposed here has been

Table 7. Mean and standard deviation of sensitivity and the specificity for EREC among conditions.

Condition	Sensitivity	Specificity
N = 150	.902 (.132)	.954 (.019)
N = 300	.963 (.077)	.956 (.017)
N = 1000	.985 (.049)	.957 (.017)
m/r = 5	.864 (.139)	.950 (.023)
m/r = 10	.982 (.045)	.951 (.014)
m/r = 20	.988 (.036)	.964 (.013)
r = 1	.952 (.103)	.933 (.024)
r = 2	.945 (.104)	.955 (.013)
r = 3	.953 (.093)	.964 (.010)
PHI = .00	.951 (.099)	.953 (.020)
PHI = .20	.948 (.100)	.960 (.013)
h2 = .30 − .40	.970 (.072)	.956 (.018)
h2 = .40 − .55	.958 (.085)	.956 (.018)
h2 = .55 − .70	.922 (.126)	.955 (.018)
Doublets = 1	.968 (.077)	.942 (.024)
Doublets = 2	.944 (.109)	.950 (.016)
Doublets = 3	.949 (.095)	.959 (.010)
Doublets = 4	.930 (.117)	.965 (.010)
Doublets = 5	.959 (.085)	.966 (.008)
Low loadings = .05	.948 (.102)	.956 (.018)
Low loadings = .10	.949 (.101)	.956 (.018)
Low loadings = .20	.953 (.095)	.956 (.018)
Size of doublets = .20 − .30	.918 (.124)	.955 (.019)
Size of doublets = .30 − .40	.962 (.086)	.956 (.018)
Size of doublets = .40 − .50	.970 (.072)	.956 (.017)

implemented in the 11.04 version of the program FACTOR (P. J. Ferrando & Lorenzo-Seva, 2017). While the researcher is allowed to select between the four diagnostic procedures discussed in this paper, MORGANA-based indices are the default option. In addition, for those users more accustomed to using R, a package is currently being developed to be allow these procedures to be used.

Discussion and conclusions

The convenience of including or not correlated residuals in FA has been, and continues to be, a controversial issue, a controversy that is more than justified, because, as discussed above, this inclusion would be potentially very prone to lead to abuses and bad practices (e.g., Bandalos, 2021). A basic result, however, is clear: if substantial correlated residuals exist and are ignored, the misspecification is expected to distort (sometimes greatly) the solution. In the EFA context, such distortions refer to incorrect assessment of the number of factors (Montoya & Edwards, 2021), and biased parameter estimates: loadings, residual variances, and, in rotated solutions, possibly also inter-factor correlations (Mulaik, 2010; Yates, 1987). So, controversies aside, the main aim of this paper: to study and propose procedures for efficiently detecting correlated residuals in EFA solutions when they are really present, seems, in our opinion, to be of clear interest.

A first interesting result obtained in this paper, is that the standard, most widespread procedure for detecting doublets is not the most efficient that can be considered. Now, at first sight, inspecting and assessing the off-diagonal elements of the residual covariance matrix seems to be the most direct approach for flagging doublets. However, this directness ignores the fact that residual correlations constrained to be zero are misspecifications that can propagate through other estimates of the model.

This result has been thoroughly discussed at least since the 1970's. However, it also seems to have been stubbornly ignored in the applied EFA literature. Oddly enough, also, assessment of doublets based on the anti-image partial correlation matrix (which is available in widespread programs, such as SPSS) appear to work clearly better than the standard approach, but it seems to have been virtually never used in applications.

In addition to assessing and comparing the two existing procedures above, we have proposed a new approach and two derived indices, which we consider to be new contributions. The basic idea of the MORGANA approach lies in the potential propagating effects of substantial doublets constrained to be zero. So, the principle of our proposal is to minimize these effects in order to obtain clearer change estimates when each doublet is or is not constrained to be zero. We admit, indeed, that our proposal is based on previous proposals. However, the results obtained here, suggests that EREC is highly efficient and outperforms procedures so far available. And ENIDE might have also value as an auxiliary index, because it focuses on the bias on the loading estimates rather than on the residuals themselves.

Following up with the contributions, for all the indices considered in the article (old and new), we have proposed procedures for obtaining efficient cutoff values, and we regard this proposal also as a new, useful contribution. An instrumental contribution, finally, is that everything we have proposed and compared here is implemented in a noncommercial widely known program, and the corresponding developments in R are quite advanced.

We indeed acknowledge that our proposal has its share of limitations and points that require further research, mainly, to carry on more extensive simulations and to undertake empirical studies to ascertain its appropriateness in practice. With the due reservations, however, we believe that we have provided the practitioner with efficient tools for detecting correlated residuals. And the fact that MORGANA detects doublets in both directions is a good starting point for future investigations that force the method into more complicated databases.

Efficiently detecting correlated residuals in EFA is only the first step in a process in which we will have to decide what to do with them. As discussed above, Thurstone (1947) proposed to "clean" the data until the doublets vanish, and so, until a classical EFA solution (2) could be correctly fitted. A second action (e.g., Mulaik, 2010; Yates, 1987) would be explicitly modeling the correlated residuals, using an extended EFA solution (1) that allows unbiased estimates to be obtained. If this second action is adopted, a careful assessment of model identification and determinacy of the solution becomes crucial, as discussed above. While our proposal implements a restriction intended to maintain the extended solution identified, the issue is far more complex and requires a careful assessment (see Hayashi & Marcoulides, 2006; Mulaik, 2010). Overall, we believe that Thurstone's recommendation is parsimonious and defensible in most cases. However, we consider our proposal here more as a basis for improved extensions of the classical EFA model.

Funding

This project has been made possible by the support of the Ministerio de Ciencia e Innovación, the Agencia Estatal de Investigación (AEI) and the European Regional Development Fund (ERDF) (PID2020-112894GB-I00).

Disclosure statement

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

ORCID

Pere J. Ferrando http://orcid.org/0000-0002-3133-5466
Ana Hernandez-Dorado http://orcid.org/0000-0001-9502-9735
Urbano Lorenzo-Seva http://orcid.org/0000-0001-5369-3099

References

Asparouhov, T., Muthén, B., & Morin, A. J. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management*, *41*, 1561–1577. https://doi.org/10.1177/0149206315591075

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438. https://doi.org/10.1080/10705510903008204

Bandalos, D. L. (2021). Item meaning and order as causes of correlated residuals in confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 1–11. https://doi.org/10.1080/10705511.2021.1916395

Blalock, J. (Ed.). (1971). *Causal models in the social sciences*. Macmillan press. https://doi.org/10.4324/9781315081663

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*, 303–316. https://doi.org/10.1177/0049124189017003004

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Butler, J. M. (1968). Descriptive factor analysis. *Multivariate Behavioral Research*, *3*(3), 355–370. https://doi.org/10.1207/s15327906mbr0303_5

Cattell, R. B. (1952). *Factor analysis: An introduction and manual for the psychologist and social scientist*. Harper.

Costner, H., & Schoenberg, R. (1973). Diagnosing indicator ills in multiple indicator models. In A. S. Goldberger, and O. D. Duncan (Eds.), *Structural equation models in the social sciences. Seminar* (Seminar press) (pp. 167–499).

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica*, *21*, 301–323. https://www.redalyc.org/pdf/169/Resumenes/Resumen_16921206_1.pdf

Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema*, *29*, 236–240. http://doi.org/10.7334/psicothema2016.304

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, *16*, 625–641. http://doi.org/10.1080/10705510903203573

Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, *10*, 524. https://doi.org/10.3389/fgene.2019.00524

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, *18*, 277–296. https://doi.org/10.1007/BF02289264

Harman, H. H. (1976). *Modern factor analysis*. University of Chicago press.

Hayashi, K., & Marcoulides, G. A. (2006). Teacher's corner: Examining identification issues in factor analysis. *Structural Equation Modeling*, *13*, 631–645. https://doi.org/10.1207/s15328007sem1304_7

Jöreskog, K. G. (2003). Factor analysis by MINRES. *To the memory of Harry Harman and Henry Kaiser*. https://www.ssicentral.com/wp-content/uploads/2020/07/lis_minres.pdf

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36. https://doi.org/10.1007/BF02291575

Kaiser, H. F. (1963). Image analysis. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 156–166). University of Wisconsin Press.

Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, *26*, 457–477. https://doi.org/10.1207/s15327906mbr2603_5

Ledermann, W. (1937). On the rank of the reduced correlational matrix in multiple-factor analysis. *Psychometrika*, *2*, 85–93. https://doi.org/10.1007/BF02288062

Lee, C. T., Zhang, G., & Edwards, M. C. (2012). Ordinary least squares estimation of parameters in exploratory factor analysis with ordinal data. *Multivariate Behavioral Research*, *47*, 314–339. https://doi.org/10.1080/00273171.2012.658340

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press. https://doi.org/10.1007/978-94-007-0753-5_1701

Lorenzo-Seva, U., & Ferrando, P. J. (2021). Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 138–147. https://doi.org/10.1080/10705511.2020.1735393

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490. http://doi.org/10.1037/0033-2909.111.3.490

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110. http://doi.org/10.1146/annurev-clinpsy-032813-153700

McDonald, R. P. (1969). A generalized common factor analysis based on residual covariance matrices of prescribed structure. *British Journal of Mathematical and Statistical Psychology*, *22*, 149–163. https://doi.org/10.1111/j.2044-8317.1969.tb00427.x

McDonald, R. P. (1985). *Factor analysis and related methods*. Psychology Press.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, *11*, 3–31. https://doi.org/10.3102/10769986011001003

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*, 413–440. https://doi.org/10.1177/0013164420942899

Mulaik, S. A. (2010). *Foundations of factor analysis* (*2nd* ed.). CRC Press. https://doi.org/10.1201/b15851

Nagy, G., Brunner, M., Lüdtke, O., & Greiff, S. (2017). Extension procedures for confirmatory factor analysis. *The Journal of Experimental Education*, *85*, 574–596. https://doi.org/10.1080/00220973.2016.1260524

Reddy, S. K. (1992). Effects of ignoring correlated measurement error in structural equation models. *Educational and Psychological Measurement*, *52*, 549–570. https://doi.org/10.1177/0013164492052003005

Revelle, W. (2021). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University, Evanston, Illinois. R package version 2.1.3. https://CRAN.R-project.org/package=psych

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

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology*, *17*, 105–129. https://doi.org/10.2307/271030

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

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371–384. https://doi.org/10.1007/BF02294623

Sörbom, D. (1975) Detection of correlated errors in longitudinal data. In: Joreskog, K. G., Sörbom, D., and Magidson, J. eds. *Advances in Factor Analysis and Structural Equation Models*, (Abt Books), pp. 171–184. https://doi.org/10.1111/j.2044-8317.1975.tb00558.x

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Thurstone, L. L. (1947). *Multiple-factor analysis; a development and expansion of the vectors of mind*. University of Chicago Press.

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

Van Kesteren, E. J., & Kievit, R. A. (2020). Exploratory factor analysis with structured residuals for brain imaging data. *BioRxiv*, *5*, 1–27 . https://doi.org/10.1101/2020.02.06.933689

Wright, S. (1968). *Genetic and biometric foundations. Evolution and the genetics of populations: A treatise in three volumes* (No. 576.58 W9301g Ej. 1 025185). The University of Chicago Press.

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. State University of New York Press.

Zhang, G., & Browne, M. W. (2006). Bootstrap fit testing, confidence intervals, and standard error estimation in the factor analysis of polychoric correlation matrices. *Behaviormetrika*, *33*, 61–74. http://doi.org/10.2333/bhmk.33.61

Zhang, L., Pan, J., Dubé, L., & Ip, E. H. (2021). blcfa: An R package for bayesian model modification in confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal 28*, 649–658. https://doi.org/10.1080/10705511.2020.1867862