CrossMark

Unrestricted factor analysis of multidimensional test items based on an objectively refined target matrix

Urbano Lorenzo-Seva[1] · **Pere J. Ferrando**[1]

Abstract

A common difficulty in the factor analysis of items designed to measure psychological constructs is that the factor structures obtained using exploratory factor analysis tend to be rejected if they are tested statistically with a confirmatory factor model. An alternative to confirmatory factor analysis is unrestricted factor analysis based on Procrustes rotation, which minimizes the distance from a target matrix proposed by the researcher. In the present article, we focus on the situation in which researchers propose a partially specified target matrix but are prepared to allow their initial target to be refined. Here we discuss RETAM as a new procedure for objectively refining target matrices. To date, it has been recommended that this kind of refinement be guided by human judgment. However, our approach is objective, because the threshold value is computed automatically (not decided on by the researcher) and there is no need to manually compute a number of factor rotations every time. The new procedure was tested in an extensive simulation study, and the results suggest that it may be a useful procedure in factor analysis applications based on incomplete measurement theory. Its feasibility in practice is illustrated with an empirical example from the personality domain. Finally, RETAM is implemented in a well-known noncommercial program for performing unrestricted factor analysis.

Keywords Partially specified target matrices · Orthogonal and oblique Procrustes rotations · Unrestricted factor analysis · Exploratory factor analysis · Confirmatory factor analysis · Multidimensional test items

Factor analysis (FA) applications designed to assess the structure of test items are frequently based on the correlated-factor model, in which item scores are assumed to measure two or more (related) dimensions. Furthermore, the pattern of the relations between the items and the factors is typically expected to approach a simple structure (Thurstone, 1947), but this expectation is generally based on not too strong or incomplete substantive measurement theory (e.g., Henson & Roberts, 2006; Myers, Jin, Ahn, Celimli, & Zopluoglu, 2015).

FA assessment of item structures can be handled using either unrestricted or exploratory factor analysis (EFA) or the more restrictive confirmatory factor analysis (CFA) model. A common difficulty in the FA of items designed to measure psychological constructs (such as personality, attitude, or psychopathology), however, is that the structures obtained using EFA tend to be rejected if they are statistically tested with a CFA model. To solve this problem, some practitioners have

proposed models based on very few items, because they seem more likely to show an acceptable fit. In addition, items are sometimes discarded ad hoc until the fit is acceptable. These procedures are likely to capitalize on chance, so they cannot be recommended (Ferrando & Lorenzo Seva, 2000).

To gain some insight into the source of the problem above, we shall first consider the unrestricted FA model based r factors, which for a given item y_j is

$$y_j = \lambda_{j1}\theta_1 + \lambda_{j2}\theta_2 + \cdots + \lambda_{jr}\theta_r + \delta_j. \quad (1)$$

In the clearest structure corresponding to this model, each item will have a salient loading on only one factor, and small or minor loadings on the remaining factors. The resulting structure will be a perfect simple structure, a more restrictive approach than the simple-structure concept advocated by Thurstone (1947), which corresponds to the idea of factorial simplicity as stated by Kaiser (1974). Now, when a factor structure of this type is assessed with CFA, the usual practice is to set the minor loadings found in the unrestricted solution (typically those below .20, .30, or even .40) to zero. The corresponding CFA model is thus given by (e.g.):

$$y_j = \lambda_{j1}\theta_1 + 0\theta_2 + \cdots + 0\theta_r + \delta_j. \quad (2)$$

✉ Urbano Lorenzo-Seva
urbano.lorenzo@urv.cat

[1] Universitat Rovira i Virgili, Tarragona, Spain

This equation corresponds to the maximum simplicity, in Kaiser's (1974) sense. In the model based on Eq. 2, it is hypothesized that the *minor* loadings found in the unrestricted solutions are consistent with exact zeros in the population: Each item is supposed to be a factorially pure measure of one sole trait, in the sense that only this trait contributes to the variance of the item (Thurstone, 1947).

Although Model 2 is regarded as the ideal model, because it assigns meaning to the estimated traits in the most unambiguous fashion (McDonald, 2005), the assumption that all the items in a multidimensional questionnaire are pure measures of a single trait is submitted to be generally unrealistic (Ferrando & Lorenzo Seva, 2000). Therefore, if Model 1 is correct for the data, and Model 2 is fitted, a bad fit would be expected due to errors of specification, which in this case would be errors of omission (significant loadings incorrectly omitted or fixed to zero).

A viable alternative for assessing item structures under incomplete measurement theory or when Model 2 is thought to be unrealistic (and so too restrictive) is to use unrestricted or exploratory FA with target or Procrustes rotation (e.g., Browne, 2001), a hybrid approach that can be conceptually situated between EFA and CFA (Asparouhov & Muthén, 2009). This approach provides an unrestricted solution in which the model parameter values (particularly zeroes) are not imposed. Rather, the sample factor solution is rotated to fit the proposed population model as closely as possible, but the model parameters are not artificially fixed to their expected values in the population.

Target rotation has been developed over more than seven decades (e.g., Mosier, 1939; Tucker, 1944), and at present many versions and approaches are derived from its basic concept. This basic concept is (a) to define a target matrix $\mathbf{H}$ in which the expected values for the loading parameters in the population are specified, and (b) to rotate the initial loading matrix $\mathbf{A}$ so as to provide a least-squares fit to $\mathbf{H}$. By way of example, consider an expected factor model with three factors ($r = 3$) and nine items ($m = 9$), in which each factor is expected to be defined by three items, and the following target matrix is proposed:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

To assess the fit of sample data to this target matrix, the unrotated loading matrix $\mathbf{A}$ should be rotated using an orthogonal Procrustes rotation (Cliff, 1966) or an oblique Procrustes

rotation (Mosier, 1939). Matrix $\mathbf{H}$ above is fully specified as it was proposed in earlier versions of the target procedure (e.g., Tucker, 1944).

A less restrictive approach, in which the proposed target is only partially specified, can also be considered, and it corresponds to latter specifications of the procedure. In this approach, only the parameter values that are expected to be zero in the population are specified in the target matrix. The partially specified target for the example above would in this case be

$$\mathbf{H}^* = \begin{pmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \\ 0 & * & 0 \\ 0 & * & 0 \\ 0 & * & 0 \\ 0 & 0 & * \\ 0 & 0 & * \\ 0 & 0 & * \end{pmatrix},$$

where the asterisks indicate the parameters that are not specified (i.e., the free parameters in the loading matrix). To assess the fit of the sample data to this partially specified target matrix, the unrotated loading matrix $\mathbf{A}$ should be rotated using an orthogonal partially specified Procrustes rotation (Browne, 1972b) or an oblique partially specified Procrustes rotation (Browne, 1972a; Gruvaeus, 1970).

In the context of fully exploratory FA (i.e., situations in which researchers do not explicitly propose a target matrix $\mathbf{H}$ or $\mathbf{H}^*$), some exploratory rotation procedures that aim to identify the best possible simple structure have also adapted the Procrustes rotation approach. These exploratory rotation methods use a two-step procedure: In the first step, a target matrix $\mathbf{H}$ (or $\mathbf{H}^*$) that maximizes factor simplicity is identified by using a simple structure rotation, and in the second step, the unrotated loading matrix $\mathbf{A}$ is rotated using a Procrustes rotation to fit the identified $\mathbf{H}$ target matrix. For example, Promax rotation (Hendrickson & White, 1964) uses Varimax rotation (Kaiser, 1958) in the first step to identify a fully specified target matrix $\mathbf{H}$, and oblique Procrustes rotation in the second step. Another example is Promin (Lorenzo-Seva, 1999): In the first step, Promin uses weighted Varimax rotation (Cureton & Mulaik, 1975) to identify a partially specified target matrix $\mathbf{H}^*$, and oblique partially specified Procrustes rotation in the second step.

In the present article, we focus on the situation in which the researcher is able to tentatively propose a partially specified target matrix $\mathbf{H}^*$ but is prepared for it to be refined. For example, in cross-cultural studies, a researcher can propose a target hypothesis for a new cultural population based on the results obtained in previous different populations. However, he or she is willing to admit that this hypothesis can be modified or refined to some extent when the new population is assessed.

This kind of situation is described by Browne (2001, p. 125), who suggested that the target might be changed after the first rotation so that any previously unspecified element in $\mathbf{H}^*$ could be specified to be zero, and new rotations could then be carried out until the researchers were satisfied with the outcome. Our proposal, however, is not guided by human judgment (as Browne suggests), but by objective criteria.

In the procedure we propose, the initial target is theoretically or substantively based, but the subsequent modifications are empirically driven. So, in the exploratory–confirmatory continuum, our proposal falls closer to the exploratory pole than the standard target rotation. This partially data-driven character means that some problems (mainly capitalization on chance) might appear, and, as we discussed below, they must be addressed. Overall, however, the results in this article suggest that our proposal is expected to be quite useful in applied research, especially when data have a complex structure. It can also be regarded as a complement to, or in some cases even a better alternative than, the analytical rotation procedures that have existed to date. From a methodological point of view, finally, the refinement strategies we propose are adapted from existing two-step rotation procedures (like Promin) that aim to identify simple structure solutions and build partially specified target matrices $\mathbf{H}^*$.

Recently, Moore, Reise, Depaoli, and Haviland (2015) proposed a procedure (iterated target rotation) in which a partially specified target rotation $\mathbf{H}^*$ is iteratively improved on the basis of an arbitrarily chosen threshold. This procedure, which was recently adapted to the context of bifactor models by Abad, Garcia-Garzon, Garrido, and Barrada (2017), bears close resemblances to our proposal. The main difference between the two is that ours starts with a substantively informed target, whereas in Moore et al.'s proposal the initial target is obtained from a standard factor rotation (i.e., it is empirically informed). Further relations and differences will be discussed below in more detail.

A new proposal for objectively refining a target matrix

In the unrestricted factor analysis of test items, a correlation matrix $\mathbf{R}$ between m items is analyzed in order to extract r factors, and the corresponding unrotated loading matrix $\mathbf{A}$ of order $m \times r$ is rotated so as to approach the proposed population model as closely as possible. In more detail, $\mathbf{R}$ is decomposed as

$$\mathbf{R} = \mathbf{P}\mathbf{\Phi}\mathbf{P}' + \mathbf{\Psi}, \tag{3}$$

where $\mathbf{P}$ is a rotated loading matrix of order $m \times r$, $\mathbf{\Phi}$ is the interfactor correlation matrix of order $r \times r$, and $\mathbf{\Psi}$ is a diagonal matrix of order $m \times m$. In the rotated loading matrix $\mathbf{P}$, the loading values describe the relationship between the m items and the r modeled factors. The partially specified target matrix $\mathbf{H}^*$ is a hypothetical proposal about how the relationships in $\mathbf{P}$ should be.

Our proposal to obtain an objectively refined target matrix (RETAM) starts from the unrotated loading matrix $\mathbf{A}$ and the partially specified target matrix $\mathbf{H}^*$ proposed by the researchers. The RETAM proposal is based on the following iterative four-step procedure:

Step 1 An initial transformation matrix $\mathbf{S}_0$ is obtained as

$$f(\mathbf{S}_0) = \text{Procrustes}\left(\mathbf{A}\mathbf{S}_0, \mathbf{H}^*\right), \tag{4}$$

where $\mathbf{S}_0$ is a transformation matrix that minimizes the distance between the product $\mathbf{B} = \mathbf{A}\mathbf{S}_0$ and the partially specified target matrix $\mathbf{H}^*$. We propose that the oblique rotation algorithm proposed by Browne (1972a) be used as the Procrustes rotation. We prefer oblique to orthogonal rotation because the former tends to produce simpler rotated matrices $\mathbf{B}$.

Step 2 In this step, the partially specified target matrix is refined. This involves comparing a threshold value with the obtained rotated loading values in $\mathbf{B}$. Where Moore et al. (2015) proposed using an arbitrary value chosen by the researchers (e.g., Moore, 2013, tested the values .05, .10, and .15), we prefer to use a more objective approach for determining the thresholds, and in particular the Promin approach (Lorenzo-Seva, 1999), which is specifically intended for a partially specified target (the scenario considered here). In Promin, a threshold value is obtained for each column of $\mathbf{B}$ in a four-step procedure. First, matrix $\mathbf{C}$ is computed as the row-normalization of $\mathbf{B}$. Second, the mean and the standard deviation of the squared elements of each column of $\mathbf{C}$ are computed. Let v $(r \times 1)$ be the vector with the means, and let s $(r \times 1)$ be the vector with the standard deviations. Third, the objective threshold value t_j $(j = 1 \ldots r)$ for each column is obtained as

$$t_j = v_j + \frac{s_j}{4}. \tag{5}$$

Finally, once the threshold values are available, a new partially specified target $\mathbf{H}_k^*$ is built: Each squared element c_{ij}^2 is compared to the corresponding threshold value t_j in order to decide whether the h_{ij} element in $\mathbf{H}_k^*$ is to be specified as a zero value or set as an unspecified parameter. Later in this article, we shall discuss the refinement strategies that can be applied to build the refined target matrix $\mathbf{H}_k^*$. When $k > 1$, if the

matrices $\mathbf{H}_k^*$ and $\mathbf{H}_{k-1}^*$ are identical (i.e., no changes have been made to the refined target matrix), then the objective refinement of the target matrix is finished, and the procedure must move on to the final step. It is acknowledged that other methods could be used to establish an objective refinement. For example, Moore (2013) proposed using the standard errors of the rotated loadings to determine whether the zero value falls within the 95% confidence intervals of the loading: If so, the loading is set to zero in the target matrix; otherwise, the element is set as a nonspecified element. Other researchers might prefer to compute the loading confidence intervals by using resampling techniques. On some occasions, the iteration could get stuck in an infinite loop (if $\mathbf{H}_k^*$ and $\mathbf{H}_{k-2}^*$ are identical but $\mathbf{H}_k^*$ and $\mathbf{H}_{k-1}^*$ are not identical): To avoid this, a maximum of number iterations can be set in advance. We must point out that in our simulation studies we never found such strange situations.

Step 3 A new transformation matrix $\mathbf{S}_k$ is obtained as

$$f(\mathbf{S}_k) = \text{Procrustes}\big(\mathbf{AS}_k, \mathbf{H}_k^*\big), \tag{6}$$

where $\mathbf{S}_k$ minimizes the distance between the product $\mathbf{B} = \mathbf{AS}_k$ and the partially specified target matrix $\mathbf{H}_k^*$. Again, it is based on oblique Procrustes rotation. Once $\mathbf{B}$ is available, go to Step 2.

Final step The final transformation matrix $\mathbf{S}$ is obtained as

$$f(\mathbf{S}) = \text{Procrustes}\big(\mathbf{AS}, \mathbf{H}_k^*\big), \tag{7}$$

where the Procrustes rotation can be either orthogonal (Browne, 1972b) or oblique (Browne, 1972a), depending on the restrictions that have been imposed on the population model. The final rotated loading matrix is obtained as

$$\mathbf{P} = \mathbf{AS}, \tag{8}$$

and the interfactor correlation matrix is obtained as

$$\mathbf{\Phi} = \mathbf{S}^{-1}\mathbf{S}^{-1'}. \tag{9}$$

Refinement strategies related to RETAM

As we already pointed out, different refinement models can be applied in Step 2 to build the refined target matrix $\mathbf{H}_k^*$. We propose three refinement strategies related to RETAM:

Make Complex (MC) The specified elements of the initial partially specified target matrix $\mathbf{H}^\bullet$ (i.e., the values defined as zero values in the target matrix) can be changed to nonspecified values in the refined target matrix $\mathbf{H}_k^*$. From a practical point of view, these would be situations in which researchers would assume that some of the items in the analysis that were initially defined as factorially pure could actually be complex items (i.e., items with cross-loadings). From a substantive point of view, the refinement of the target matrix is the least possible refinement, and the refined target matrix $\mathbf{H}_k^*$ does not importantly contradict the substantive model on which the initial partially specified target matrix $\mathbf{H}^\bullet$ is based.

Make Simple (MS) The nonspecified elements of the initial partially specified target matrix $\mathbf{H}^\bullet$ can be changed to specified values in the refined target matrix $\mathbf{H}_k^*$ (i.e., values defined as zero values in the target matrix). This is the refinement proposed by Browne (2001, p. 125). From a practical point of view, these are the situations in which researchers assume that the items in the analysis might be incorrectly assigned to a factor in the initial target matrix (i.e., misspecified items). From a substantive point of view, the refinement of the target matrix is more important than in MC, and the refined target matrix $\mathbf{H}_k^*$ can importantly contradict the substantive model on which the initial partially specified target matrix $\mathbf{H}^\bullet$ is based. Researchers must carefully study how the refinement procedure has changed the initial proposal of the target matrix $\mathbf{H}^\bullet$ in order to assess whether the final target matrix $\mathbf{H}_k^*$ can be accepted from a substantive point of view.

Complete Refinement (CR) All elements of the initial partially specified target matrix $\mathbf{H}^\bullet$ can change their role in the partially specified target matrix. From a practical point of view, these are situations in which researchers assume that (1) the items in the analysis can be complex items (i.e., items with cross-loadings), and (2) the items in the analysis can be incorrectly assigned to a factor in the initial target matrix (i.e., misspecified items). From a substantive point of view, the refinement of the target matrix is more important than in either of the previous strategies, and the refined target matrix $\mathbf{H}_k^*$ can strongly contradict the substantive model on which the initial partially specified target matrix $\mathbf{H}^\bullet$ is based. Once more, researchers must carefully study how the refinement procedure has changed the initial proposal of the target matrix $\mathbf{H}^\bullet$ in order to assess whether the final target matrix $\mathbf{H}_k^*$ can be accepted from a substantive point of view. In addition, this strategy can be defined as the most exploratory refinement model.

The performance of RETAM in different situations will be assessed below, to determine whether any of the refinement strategies is superior to the others.

Some substantive and practical considerations

As we discussed above, when the CFA model is used in an exploratory way (e.g., Browne, 2001) by modifying and discarding items ad hoc until the fit is acceptable, the problem of capitalization on chance is likely to occur. In our view, modifications of an initial CFA solution can only be acceptable if the changes are sound and in agreement with the theory, and if the problem of capitalizing on change is satisfactorily addressed.

Because RETAM starts from a theoretically derived target and then uses empirically derived modifications, our procedure can be accused of the same problems mentioned above in the CFA context. For this reason, we suggest that researchers use RETAM in the following way: First, the original sample must be split into two random halves; second, the RETAM procedure should be applied to the first subsample to obtain a refined target matrix; and third, the refined target matrix should be taken as a fixed target matrix (without further refinements) to be used in the second subsample. If the rotated loading matrix in the second subsample is congruent with the rotated loading matrix in the first subsample, then researchers will be more confident that the final solution has not merely been specifically fitted to the sample data, but that it generalizes to the population for which the analysis is intended.

The cross-validation study requires that the sample be divided into two subsamples. As a consequence, the sample may need to be larger than the sample size needed when computing a standard EFA. In the next section, we shall assess which sample sizes would be advisable when using the MC, MS, and CR strategies.

A second potential problem when RETAM is used concerns the order and sign indeterminacies of the target and rotated pattern matrices, in the sense that the order of the factor columns is interchangeable, and each column is interchangeable with its negative (e.g., Myers, Ahn, Lu, Celimli, & Zopluoglu, 2017). In a real application, particularly when the procedure is based on an initial target specification, like the one here, this problem is expected to be unimportant. However, it potentially exists, so we must recommend that researchers control the process and use appropriate reordering or sign-change modifications, should they be needed.

Simulation study

The simulation studies reported in this section were intended to assess the functioning of RETAM under different scenarios. In general terms, the design attempted to mimic the conditions expected in empirical applications, and so to provide realistic choices. The main settings in our simulation study were based on the simulation studies by Myers, Ahn, and Jin (2013), and Myers, Jin, Ahn, Celimli, and Zopluoglu (2015). Two main

preliminary hypotheses can be advanced from these simulation studies when no refinement is applied. First, factorial congruence (to be defined below) is expected to increase with the number of targets. Second, there is an interactive effect between the number of targets and communality, so that the increase in congruence with the number of targets decreases when communality is high. As for the specific performance of the proposed RETAM strategies, we preferred not to advance any hypothesis and to maintain the study as essentially exploratory.

Method

The design consisted of two simulation studies. Study 1 explored the capabilities of the MC, MS, and CR strategies associated with RETAM to recover the population loading matrix. Study 2 explored the sample sizes needed to carry out the cross-validation analysis.

We specified three population models, each with a different level of communality (low, wide, or high). The models were taken from the population loading matrices proposed by MacCallum, Widaman, Preacher, and Hong (2001), which included 20 measured variables ($m = 20$) and three factors ($r = 3$). We selected these population matrices because, as Myers et al. (2015) pointed out, they have characteristics that mimic realistic situations in the context of EFA and have already proved to be useful in a number of simulation studies in the literature. For each of the three population models, we built 12 partially specified target matrices in which the number of specified elements in the target matrices ranged from 12 to 30. The specified values in the target matrix were set to zero (i.e., they were expected to be zero values in the population loading matrix). In addition to changing the number of specified values, we also changed the precision of the specification in the target matrix, in order to introduce some level of error. We considered that loading values lower than absolute .20 in the population model should be set as specified values in the target matrix (i.e., values expected to be zero in the population), and as unspecified values in the target matrix otherwise. If we consider this criterion, the numbers of unspecified values in the target matrices should be 29, 30, and 32, respectively, for the population models with low, wide, and high communality. On the other hand, two kinds of error could be produced:

1. If the loading value in the population model is lower than absolute .20 and is set as an unspecified value in the target matrix, then an error has been committed. We call this type of error *Free-errors*, because the element has been erroneously set as a free element in the target matrix.
2. If the loading value in the population model is larger than absolute .20 and is set as a specified value in the target matrix, then an error has also been committed. We call this type of error *Fixed-errors*, because the element has been erroneously set as a fixed element in the target matrix.

Table 1 summarizes the percentages of Free-errors and Fixed-errors for the partially specified target matrix related to each population matrix. For example, for the population model with wide communality, we constructed three target matrices with 30 unspecified elements (i.e., free elements) and 30 specified elements (i.e., fixed elements). In the first target matrix, no error was introduced (i.e., the free and the fixed elements in the target matrix were all properly defined). In the second target matrix, 27% of the free elements (i.e., a total of eight elements out of 30) were Free-errors, and 27% of the fixed elements (i.e., a total of eight elements out of 30) were Fixed-errors. In the third target matrix, 50% of the free elements (i.e., a total of 15 elements out of 30) were Free-errors, and 50% of the fixed elements (i.e., a total of 15 elements out of 30) were Fixed-errors. As a second example, again for the population model with wide communality, we constructed three target matrices with 36 unspecified elements (i.e., free elements) and 24 specified elements (i.e., fixed elements). In the first target matrix, 20% of the free elements (i.e., a total of six elements out of 30) were Free-errors, and no error was introduced in the fixed elements. In the second target matrix, 40% of the free elements (i.e., a total of 12 elements out of 30) were Free-errors, and 20% of the fixed elements (i.e., a total of six elements out of 30) were Fixed-errors. In the third target matrix, 60% of the free elements (i.e., a total of 18 elements out of 30) were Free-errors, and 40% of the fixed elements (i.e., a total of 12 elements out of 30) were Fixed-errors. This procedure was followed to construct the 36 partially specified target matrices (i.e., 12 target matrices for each population matrix). The 36 partially specified target matrices were checked to confirm that the rotation identification conditions were met (see Myers et al., 2017; Myers et al., 2015). To help other researchers replicate our study, we can offer

interested readers the set of target matrices that we produced. As an independent variable of the simulation studies, we included the levels of Free-errors and Fixed-errors in the partially specified target matrix used to rotate each sample loading matrix. The number of specified elements in the target matrix (12, 18, 24, or 30) was also recorded.

Overdetermination ($m : r = 20 : 3$) and model error (i.e., population RMSEA = .065) were kept constant. The level of model error has been defined in the literature as a *fair fit* (Browne & Cudeck, 1992). Interfactor correlations were not manipulated, either. Moore et al. (2015) carried out a simulation study on target rotations, and they reported that the correlations among the factors in the population had little to no influence on the relative abilities of the rotations to approximate the population factor structure. In our simulation study, we set the interfactor correlations to zero in the population.

Study 1

Manipulated factors

The study was based on a $3 \times 3 \times 4$ design and 500 replicas per condition. The independent variables were (1) sample size: $N =$ 100, 300, 500; (2) communality: low (item communalities between .20 and .40, with an average of .32), wide (item communalities between .20 and .80, with an average of .49), and high (item communalities between .60 and .80, with an average of .69); and (3) number of specified targets: 12, 18, 24, 30. Please note that the 36 partially specified target matrices were used in the study, and that the levels of Free-errors and Fixed-errors in the matrices used to rotate each sample loading matrix were also recorded as independent variables in the study.

Table 1 Levels of specification error for the 36 targets: 12 targets related to low communality, 12 targets related to wide communality, and 12 targets related to high communality

Communality	Target elements		Error in target matrix					
	Free	Fixed to Zero	Free	Fixed	Free	Fixed	Free	Fixed
Low	30	30	3%	0%	29%	28%	52%	52%
	36	24	23%	0%	42%	21%	61%	41%
	42	18	42%	0%	58%	17%	71%	31%
	48	12	61%	0%	71%	10%	81%	21%
Wide	30	30	0%	0%	27%	27%	50%	50%
	36	24	20%	0%	40%	20%	60%	40%
	42	18	40%	0%	57%	17%	70%	30%
	48	12	60%	0%	70%	10%	80%	20%
High	30	30	0%	6%	21%	25%	46%	47%
	36	24	14%	0%	36%	19%	57%	38%
	42	18	36%	0%	54%	16%	68%	28%
	48	12	57%	0%	68%	9%	79%	19%

Rotation identification

Although the population loading matrices provided by MacCallum et al. (2001) were not rotated (because the authors provided them already rotated using direct quartimin rotation), the sample loading matrices were rotated using oblique partially specified rotation (Browne, 1972a), in which the partially specified target matrices were the 36 matrices summarized in Table 1. As in Myers et al. (2015), we checked that the conditions for factor specification were met for each of the 36 target matrices.

Data generation

The simulated data were generated by a linear common factor model, which included both major and minor factors. The minor factors aimed to be a realistic representation of empirical cases (MacCallum & Tucker, 1991). Because a common factor model with a limited number of common factors will never fit exactly at the population level, each variable is considered to be composed of one part that is consistent with the common factor model and another that is not. The latter is called the *model error* and is represented by the minor factors (based on the middle model by Tucker, Koopman, & Linn, 1969). This approach has been taken in earlier research to assess the performance of PA (see, e.g., MacCallum & Tucker, 1991). In the simulation study, the sample correlation matrices were modeled as

$$\mathbf{R} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Theta}^2 + \mathbf{\Delta}_{ME} + \mathbf{\Delta}_{SE}, \tag{10}$$

where $\mathbf{\Lambda}$ is the population loading matrix, $\mathbf{\Theta}$ is a diagonal matrix of unique coefficients, $\mathbf{\Delta}_{ME}$ is the model error in the covariance structure, and $\mathbf{\Delta}_{SE}$ is the sampling error. As population loading matrices and unique coefficients, we used the data offered by MacCallum, Widaman, Zhang, and Hong (1999) related to 20 observed variables, three common factors, and the three different levels of communality. The model error was manipulated so that a population RMSEA = .065 was expected. The sample error was manipulated by using samples of different sizes (100, 300, 500) drawn from a normal random distribution (0, 1). In the first step, a correlation matrix $\mathbf{R}^*$ was obtained as $\mathbf{R}^* = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Theta}^2 + \mathbf{\Delta}_{ME}$. Then we computed the Cholesky decomposition of $\mathbf{R}^* = \mathbf{L}'\mathbf{L}$, where $\mathbf{L}$ is an upper triangular matrix. The sample data matrix of continuous variables $\mathbf{X}$ was finally obtained as $\mathbf{X} = \mathbf{ZL}$, where $\mathbf{Z}$ is a matrix of random standard normal scores, with rows equal to the corresponding sample size and a number of columns equal to the corresponding number of variables.

Dependent variables

The population and the sample loading matrices were compared. Please note that the sample loading matrices were rotated with no target refinement and using MC, MS, and CR refinement strategies. Congruence and discrepancy indices were used to assess the degree to which the true generated structures were recovered. The congruence index was the Burt–Tucker coefficient of congruence, a measure of profile similarity (see Lorenzo-Seva & ten Berge, 2006) that is defined as

$$\phi(x,y) = \frac{\sum x_i y_i}{\sum x_i^2 \sum y_i^2}. \tag{10}$$

Equation 10 was used to assess the congruence between the columns of the population loading matrix and the columns of the fitted loading matrices. The overall congruence between two loading matrices is usually reported by calculating the average of the column congruence. Lorenzo-Seva and ten Berge (2006) pointed out that a value in the range [.85–.94] corresponds to a fair similarity, whereas a value higher than .95 implies that the factor solutions compared can be considered equal. The discrepancy index was the root-mean squared residual (RMSR) between the population model and the data-fitted model, a measure of profile distance that is defined as

$$\text{RMSR}(\mathbf{X}, \mathbf{Y}) = \sqrt{(1/mr) \sum_i^m \sum_j^r \left(x_{ij} - y_{ij}\right)^2}. \tag{11}$$

To analyze the size of the effects, analyses of variance were carried out with the IBM SPSS Statistics version 20 program. Cohen (1988, pp. 413–414) suggested that for eta-squared (η^2) effect sizes, threshold values of .02 represent small effects, .13 medium effects, and .26 or more large effects.

Study 2

This simulation study explored the sample sizes needed to perform the cross-validation analysis. Mainly it was a replication of the previous simulation study, except for the way the sample size was manipulated. The idea was to generate a sample of very low size ($N = 50$), randomly divide the sample in order to refine the target using the first random sample, and rotate the loading matrices obtained in both (the first and the second) samples. If the congruence value between the two rotated solutions was lower than .96, then a new sample would be obtained with a sample size of ten extra observations. When the threshold value of .96 was obtained, we recorded the sample size of the final sample. The RETAM was applied with the MC, MS, and CR strategies. For each experimental condition, 500 replications were computed. To determine an advisable threshold value for the size of the sample needed

when each refinement strategy was computed, the distributions of the sample sizes of the two conditions in the study (i.e., the *Number of targets specified as zero* and *Communality*) were recorded for the 500 replications, and the 95th percentile was taken as the advisable threshold.

Results

Study 1 The correlations of the percentages of Free-errors and Fixed-errors in the target matrices with the congruence and discrepancy indices are shown in Table 2. When no refinement was applied, as the percentage of both types of error increased, the congruence index decreased and the discrepancy index increased. This effect was more important for Free-errors (i.e., when the loading value in the population model was lower than absolute .20 and was set as an unspecified value in the target matrix). Furthermore, performance was observed to depend on the refinement strategy used:

1. When the MC strategy was used, congruence and discrepancy values were independent of the percentage of Free-errors;
2. When the MS strategy was used, congruence and discrepancy values were independent of the percentage of Fixed-errors (when the loading value in the population model was larger than absolute .20 and was set as a specified value in the target matrix);
3. When the CR strategy was used, congruence and discrepancy values were independent of both percentages of error.

Tables 3 (congruence results) and 4 (discrepancy results) summarize the behavior of the RETAM approach. When no refinement was computed, the population loading matrix was systematically recovered badly (i.e., congruence values were lower than .95, and discrepancy indices were larger than .10). The MC refinement strategy correctly recovered the population solution when the sample was large, the communality was high, and the number of elements in the target fixed as zero was low. The MS refinement strategy performed similarly, except for the number of elements fixed to zero: Now the more elements were fixed, the better the population matrix was recovered. The CR strategy systematically recovered the population loading matrix. In terms of discrepancy, low sample sizes and wide communality were the most difficult situations to manage.

Table 5 summarizes the sizes of the main effects in the simulation study and shows the interactions that produced effect sizes larger than .02 in terms of η^2. When no refinement was computed, the main effects of communality and number of targets specified as zero (and their interaction) were at some point substantial, a result that agrees with the preliminary hypotheses above. When MC refinement was applied, only the main effect of communality was substantial; this means that when the communality is low (i.e., low loading values are observed in the loading matrix), MC refinement may have trouble recovering the population loading matrix. When MS refinement was applied, only the main effect of number of targets specified as zero was substantial; this means that when just a few elements in the target matrix are fixed to zero, the MS refinement may help recover the population loading matrix. Finally, when the CR strategy was applied, the main effects of sample size and communality showed substantial effects; this means that this approach works better with large samples and high communality.

Study 2 Table 6 shows the 95th percentiles for each condition in the study after the 500 replications. When the MC refinement is used, the largest samples are advisable if the number of elements specified as zero is high and the communality low. In our study, a sample of 420 observations would be needed in these conditions. On the other hand, when the number of elements specified as zero is low and the communality high, a sample of 80 observations could suffice. When the MS refinement is used, the largest samples are advisable if the number of elements specified as zero and the communality are both low. In our study, a sample of 460 observations would be needed in these conditions. Again, when the number of elements specified as zero and the communality are both high, a sample of 80 observations could suffice. When the CR strategy is used, the number of elements specified as zero makes

Table 2 Correlations of congruence and discrepancy values with the percentages of error in the target matrices

Refinement strategy	Congruence index		Discrepancy index	
	Error in free elements	Error in fixed elements	Error in free elements	Error in fixed elements
No refinement	**− .472**	**− .333**	**.426**	**.202**
Make complex	− .093	**− .288**	.121	**.300**
Make simple	**− .415**	− .090	**.364**	.029
Complete refinement	− .043	− .064	.047	.046

Correlation values larger than .20 are printed in bold

Table 3 Averages and standard deviations (given in parentheses) of the congruence index

Condition	No refinement	Make complex	Make simple	Complete refinement
Overall	.894	.948	.942	**.970**
	(.087)	(.057)	(.067)	(.029)
Sample size				
$N = 100$	.884	.934	.928	**.956**
	(.085)	(.060)	(.069)	(.039)
$N = 300$	.898	.953	.948	**.976**
	(.087)	(.054)	(.064)	(.020)
$N = 500$	.901	**.953**	**.950**	**.979**
	(.087)	(.054)	(.066)	(.017)
Communality				
High	.907	**.985**	**.966**	**.992**
	(.085)	(.021)	(.057)	(.006)
Wide	.859	.926	.912	**.965**
	(.100)	(.064)	(.083)	(.017)
Low	.916	.934	.948	**.953**
	(.058)	(.056)	(.044)	(.038)
Number of targets specified as zero elements				
12	.870	**.959**	.904	**.971**
	(.107)	(.041)	(.089)	(.026)
18	.873	**.950**	.937	**.971**
	(.092)	(.060)	(.067)	(.026)
24	.923	**.953**	**.965**	**.970**
	(.056)	(.050)	(.036)	(.029)
30	.911	.934	**.961**	**.968**
	(.071)	(.069)	(.043)	(.035)

Values higher than the threshold of .95 are printed in bold

no difference, and only communality need be taken into account. When the communality is low, a sample size of 390 might be advisable, whereas a sample of 80 observations may be enough when the communality is high.

Finally, as was pointed out by an anonymous reviewer, it should be noted that the cross-validation procedure is only useful for analyzing consistency across half samples, but not necessarily for drawing conclusions about accuracy.

Illustrative examples with real data

A 38-item version of the Overall Personality Assessment Scale (OPERAS; Vigil-Colet, Morales-Vives, Camps, Tous, & Lorenzo-Seva, 2013) was administered to a sample of 4,085 participants. The scales aim to assess six independent factors: extraversion (EX; seven items), emotional stability (ES; seven items), conscientiousness (CO; seven items), amiability (AM; seven items), openness to experience (OE; seven items), and social desirability (SD; three items). All 38 items are positively worded and use a 5-point Likert response format.

Examination of the item scores showed that the response distributions were generally skewed. So the item scores were treated as ordered-categorical variables, and the factor analysis based on the polychoric interitem correlations was the model chosen to fit the data. This model is an alternative parameterization of the multidimensional item response theory graded response model.

Since the interitem polychoric correlation matrix had good sampling adequacy (KMO = .871), six factors were extracted by using robust factor analysis based on the diagonally weighted least squares criterion, as implemented in the FACTOR program (Ferrando & Lorenzo-Seva, 2017), and these reached acceptable goodness-of-fit levels: RMSEA = .036 (values between .010 and .050 are considered to be close), CFI = .970, GFI = .989, and WRMR = 0.026.

Because each item on the scale was expected to be related to a single factor, a rotation target might easily be proposed by a researcher: A partially specified target matrix was proposed in which each item had a nonspecified value in the factor that it was expected to assess, and zeros otherwise. This target matrix indicates that each item was expected to be a good

Table 4 Averages and standard deviations (given in parentheses) of the discrepancy index

Condition	No refinement	Make complex	Make simple	Complete refinement
Overall	**.189**	**.116**	**.136**	.092
	(.129)	(.054)	(.117)	(.032)
Sample size				
$N = 100$	**.199**	**.132**	**.154**	**.111**
	(.119)	(.051)	(.131)	(.034)
$N = 300$	**.185**	**.110**	**.129**	.086
	(.130)	(.052)	(.112)	(.027)
$N = 500$	**.183**	**.105**	**.125**	.081
	(.136)	(.053)	(.104)	(.027)
Communality				
High	**.220**	.080	**.121**	.061
	(.179)	(.033)	(.135)	(.011)
Wide	**.220**	**.157**	**.184**	**.121**
	(.101)	(.054)	(.107)	(.016)
Low	**.128**	**.111**	**.103**	.096
	(.042)	(.040)	(.107)	(.030)
Number of targets specified as zero elements				
12	**.248**	**.107**	**.205**	.092
	(.202)	(.058)	(.194)	(.032)
18	**.201**	**.117**	**.137**	.092
	(.101)	(.058)	(.071)	(.032)
24	**.150**	**.111**	.099	.094
	(.067)	(.049)	(.047)	(.034)
30	**.157**	**.128**	**.102**	.093
	(.068)	(.058)	(.049)	(.033)

Values larger than the threshold of .10 are printed in bold

indicator of a single factor (i.e., to have a single salient loading on a factor). However, some researchers believe that this expectation is not realistic, and that personality items are frequently complex indicators (see, e.g., Woods & Anderson, 2016). The complexity of personality items is defined in the context of the periodic table of personality. In this context, the largest salient loading of a personality item informs as to the factor that this item mainly assesses, whereas secondary salient loadings (i.e., loading values of the item that are not as large as the main salient loading but still large enough to be

Table 5 Univariate analysis of variance effect sizes η^2

Effect	No refinement		Make complex		Make simple		Complete refinement	
	C	D	C	D	C	D	C	D
Main effects								
Sample size (N)	.01	.00	.03	.05	.02	.01	.12	**.16**
Communality (h)	.08	.11	**.21**	**.35**	.11	.09	**.31**	**.58**
Number of targets specified as zero elements (T)	.07	.09	.03	.02	**.13**	**.13**	.00	.00
Two-way interactions								
N × h	.01	.00	.02	.01	.02	.01	.08	.04
h × T	.07	.08	.03	.02	.07	.06	.00	.00

Only substantial interactions are printed. C stands for the congruence index, and D stands for the discrepancy index. Values of at least the threshold of .13 are printed in bold.

Table 6 Sample sizes recommended to compute the cross-validation study

Refinement strategy:	Make complex			Make simple			Complete refinement		
Communality:	Low	Wide	High	Low	Wide	High	Low	Wide	High
Number of targets specified as zero elements									
12	390	170	80	460	350	140	390	160	80
18	390	170	80	430	200	100	390	160	80
24	400	170	90	410	180	90	390	160	80
30	420	180	90	400	179	80	390	160	80

meaningfully interpreted) define other factors in which a person's response to the item also gives some substantial information. Although it is easy to propose the main loading of an item in advance, it is not so easy to propose secondary salient loadings. In conclusion, although the researcher can easily propose a partially specified target, he or she might also expect some items not to be pure indicators of a single factor, and must be prepared to accept that some items could turn out to be complex indicators.

In addition to substantive dimensions, OPERAS also aims to measure SD. Now, because responses to personality items are frequently expected to be biased by SD (see, e.g., Ferrando, Lorenzo-Seva, & Chico, 2009), it is reasonable to assume that some of the items analyzed here were complex, with a main salient loading on the corresponding personality factor and a secondary loading on the SD factor.

In summary, then, this is a research context in which an initial target hypothesis can be proposed for all of the items under study. At the same time, however, it is also reasonable to consider that this hypothesis could be modified or refined to a certain extent. This midpoint location between exploratory and confirmatory is a perfect scenario to illustrate how RETAM can be useful to practitioners, and to this end, three different approaches will be presented. In each approach, the researcher adopts a different attitude to the dataset. Finally, since the five personality factors are typically considered orthogonal in the literature, the rotations computed were systematically orthogonal rotations. To help other researchers best understand our results, we can offer interested readers the set of targets and rotated loading matrices that we obtained in the three analyses that follow.

First analysis

In the first analysis, the aim was to propose an initial hypothesis that assumed the simplest factor solution (i.e., that each item was related to a single factor). Although the researcher feels confident that he has correctly identified which factor is related to each item, other substantial secondary loadings can also be expected. As a consequence, the most advisable

RETAM strategy would be to allow the target to become more complex than initially proposed (i.e., the MC refinement strategy).

RETAM made eight changes to the target. Items 33 and 34 (which were expected to assess OE) were adjusted so that they could show a much more complex behavior: They were also expected to load on the AM personality factor and to the SD factor. As an example, the content of Item 33 is "I feel curious about the world around me." Three other personality items were also expected to become complex items and to load on another personality factor and on the SD factor: Item 18 (CO), Item 27 (AM), and Item 28 (AM). As an example, the content of Item 27 is "I am very critical of others." Finally, an SD item (37) was allowed to show a salient loading on a personality factor (AM). The content of this item is "Sometimes I have taken advantage of someone." As can be observed, the items related to CO and AM are the most susceptible to becoming biased by SD. At the same time, items related to SD can also be biased by some personality factor (like AM).

We compared the rotated pattern matrix obtained (1) when the researcher-defined target matrix was used and (2) when the RETAM refined target matrix was used. Both rotated pattern matrices were quite similar: The congruence indices between the corresponding columns ranged from .985 to .995. These values were clearly larger than the threshold of .95 (Lorenzo-Seva, & ten Berge, 2006). Secondary loading values in the rotated pattern related to the researcher-defined target seemed to suggest that some of the items were not as simple as the ones proposed in the target matrix. However, these secondary loadings were largest in the rotated pattern related to the RETAM refined target matrix. In this regard, the rotation based on the refined target helped to better understand the complexity of some items.

It is also interesting to note that some items, which seemed simple in the rotation based on the researcher-defined target matrix, showed their complexity in the rotation based on the RETAM refined target matrix. An example is Item 1 ("I make friends easily"). This item was expected to be related only to the EX factor. However, the rotation based on the RETAM refined target matrix

suggested that it is actually a complex item that is also related to the AM factor.

Second analysis

In the second analysis, the simplest initial hypothesis was again proposed, as above. However, now the researcher does not feel so confident of correctly identifying which factor is related to each item, which means that some items could have a single salient loading on an unexpected factor. In addition, the researcher could expect to observe some substantial secondary loadings. As a consequence, the most advisable RETAM strategy here would be to allow the target to be fully refined (i.e., the CR strategy).

A total of 13 changes were made to the target: Ten values specified as zero by the researcher were set to be nonspecified values, and three values set to be nonspecified values by the researcher were set to be specified zero values. Overall, the most remarkable change was that the three items (Items 26, 27, and 28) that the researcher expected to define the AM factor were changed to become items that were expected to load on the same factor as the SD items. In addition, seven items were expected to show secondary salient loadings. The table also shows the rotated pattern matrix based on the refined target. In this context, the factor that the researcher expected to be related to SD turned out to be a mixture of SD and AM. In addition, the factor that the researcher expected to be related to AM was a mixture of AM and OE.

Third analysis

In the third analysis, the researcher felt confident enough to propose six items, each of which was expected to be a good indicator of a single different factor (i.e., to have a single salient loading on a factor). At the same time, even if she expected the items to be simple indicators of a single factor, she preferred not to propose a hypothesis for the other items in the analysis. It must be noted that this is a weak target matrix (since very few values were defined). As a consequence, the most advisable RETAM strategy would be to allow the target to become simpler than initially proposed (i.e., the MS refinement strategy).

A total of 156 changes were made to the target. To summarize these changes, 26 items were defined by the refined procedure as simple indicators of a single factor, and six items (Items 6, 7, 20, 32, 33, and 34) were defined as complex items (with two salient loading values). In addition, once again the three items (Items 26, 27, and 28) that the researcher expected to define the AM factor loaded on the same factor as the SD items. Both rotated pattern matrices were quite similar: the congruence indices between corresponding columns ranged from .961 to

.995. However, the loading simplicity index (Lorenzo-Seva, 2003) reported that the simplicity of the rotated pattern based on the refined target matrix was larger (value of .417) than the simplicity of the rotated pattern based on the researcher-defined target (value of .397).

Comparison of the three analyses

To determine whether the three analyses based on the different refinement strategies produced substantially different rotated pattern matrices, we computed the congruence coefficients between the columns of the rotated pattern matrices. As can be observed in Table 7, the outcomes related to four factors (EX, EE, CO, and OE) remained quite constant, regardless of the refinement strategy. However, the strategy based on MC produced slight differences in the outcomes related to the AM factor (congruence values equal or slightly larger than .95). The differences were clearer in the SD factor.

It is interesting to point out that two different refinement strategies (MS and CR) led to quite congruent rotated pattern solutions. This was due more to the initial target proposed by the researcher than to the procedure itself. If the MS approach had been used in this dataset based on the first researcher-defined target matrix, then the MC and MS strategies would have been much more congruent with each other. This means that the usefulness of each strategy to a researcher would depend on the researcher-defined target, and his or her personal position when analyzing each particular dataset. It must also be noted that if a set of items have a clear and strong relationship with one another (like the items in EE in our illustrative example), the same final result is expected to be attained, regardless of the position of the researcher or the chosen refinement target procedure. However, if the items have a more ambiguous relationship, then researchers must decide on their personal position and the refinement target they will use.

Table 7 Congruence coefficient value among the three RETAM strategies used in the three analyses of the illustrative example

RETAM strategies	EX	EE	CO	AM	OE	SD
MC vs. CR	.995	.997	.992	.954	.996	**.940**
MC vs. MS	.995	.998	.988	.950	.995	**.920**
CR vs. MS	.998	.998	.995	.999	.997	.994

EX, extraversion; ES, emotional stability; CO, conscientiousness; AM, amiability; OE, openness to experience; SD, social desirability. Values lower than the threshold of .95 are printed in bold

Conclusions

The outcomes of the simulation study seem to suggest that no one RETAM refinement model can be regarded as systematically superior to the others. When the researcher can set a large number of elements to zero in the target matrix, the MC refinement is advisable. When the researcher prefers to set a low number of elements to zero in the target matrix, then the MS refinement is advisable. Finally, the CR strategy may be useful to researchers who can modify their initial target matrix (and possibly the implicit substantive model) during the analysis. This would be the case for a factor analysis that is closer to a pure EFA. Researchers, however, must be aware that this exploratory approach may require a larger sample (especially if communalities are low).

Discussion

Traditionally, factor analysis has been artificially split into two approaches: exploratory factor analysis (to be applied when the researcher does not have a hypothesis about the population model) and confirmatory factor analysis (to be applied when the researcher has a definitive hypothesis about the population model). In the first situation, exploratory rotations can be used. In the second situation, the researcher assesses the fit of the proposed model using a sample (and probably working with specific software, such as LISREL or Mplus). However, between these two extremes, a large number of situations are sometimes closer to one of the two poles and sometimes are in the middle (e.g., Henson & Roberts, 2006; Myers et al., 2015). Applied researchers who find themselves in one of these indefinite intermediate points have no clear methodological alternative. They must choose between (1) dropping their weak tentative hypothesis (and computing a full exploratory analysis) or (2) making as if the tentative hypothesis was a definitive one and proceeding with a CFA. Our proposal is aimed at researchers who have a tentative hypothesis and are prepared to refine this hypothesis in an exploratory way (e.g., Browne, 2001).

We have proposed RETAM as a new procedure for objectively refining target matrices in the context of unrestricted FA. To date, it has been recommended that this kind of refinement be guided by human judgment (see Browne, 2001). Furthermore, the approach has already being used by applied researchers. For example, Ayr, Yeates, Taylor, and Browne (2009) presented an FA of postconcussive symptoms in children with mild traumatic brain injuries. They made a number of Procrustes rotations based on progressively refined target matrices, used a threshold value of .40 to correct misspecifications of the items to the factors, and finished with a multidimensional test of 39 items and four factors. Conceptually, their approach was similar to our refinement

Model 2. In comparison with these proposals, however, ours (a) is more objective (because the researcher does not need to set a necessarily arbitrary threshold value), (b) is computed automatically (there is no need to manually compute a number of factor rotations every time), and (c) controls for capitalization on chance, as we discuss below. Methodologically, our approach incorporates proposals that already exist in the factor-analytic literature, mainly the iteration procedure for refining a target (Moore et al., 2015) and the Promin procedure (Lorenzo-Seva, 1999) for objectively defining the thresholds. The resulting proposal in which these developments are combined, however, appears to be new.

Overall, RETAM can be regarded as a hybrid procedure, in the sense that it combines theoretically derived specifications (the initial target) with analytical, empirically informed specifications (the refinement procedure). In this way, our procedure falls between a pure analytical rotation and a target rotation. In our view, it is of particular interest for those applications in which a priori knowledge or theory allow for a specification that is more detailed than merely setting the number of factors and deciding whether or not they are correlated (i.e., pure analytical rotation). However, it is not so complete as to allow a definite target to be specified. Rather, the initial target might be too strict (i.e., complex items wrongly specified as pure items) or too lenient, and in both cases the proposed procedure is expected to be able to arrive at a more correct final target.

As long as the initial information available is more than is required by a purely analytical rotation, we believe that our proposal is more appropriate. A wide variety of analytical rotation options exist, which understand the structure of item–factor relations in different ways. So, if an initial structure is proposed (albeit only tentatively), the process becomes less determined by the analytical simplification criterion (i.e., rows, columns, or both) and more guided by theory. To understand this point in more detail, note that RETAM does not seek to simplify rows or columns as analytical criteria do, but rather to identify a pattern of salient loadings, starting from a theoretically informed initial pattern.

Because different analytical rotation criteria tend to lead to diverging solutions when the factor structure is complex (e.g., Moore et al., 2015), we believe that the extra initial guidance provided by RETAM will be especially important for complex structures, and the results of our simulation point in this direction. Despite this, however, the simulation results do not suggest a consistent superiority of the strategies related to their use. Some researchers prefer to specify the initial target by using very few items per factor (i.e., markers), and in this case, MS seems to be the best refinement. Others tend to start with highly restrictive targets (i.e., following a typical CFA approach), and if they do, MC is the strategy to choose. CR is the strategy closest to pure EFA, so if would be justified in the case of a very weak measurement theory.

Because of its partially data-driven component, RETAM is potentially prone to theoretically unjustified ad hoc modifications, capitalization on chance, and problems of pattern indeterminacy. Thus, we strongly emphasize that the potential problems should be addressed by using a well-designed cross-validation schema (Simulation Study 2 provided guidance on this point) and carefully controlling the process, which includes a check that the final rotated loading matrix has a substantive interpretation that is consistent with the theory.

Our illustrative example aimed to show how a researcher can advance a hypothesis based on the factor in which (a) each item is expected to be a good indicator of a single factor, but (b) secondary factor loadings can be expected (which are much more difficult to predict). In addition, we showed that, depending on their personal positions, researchers can propose (a) a strong hypothesis, or (b) a weak hypothesis. Both of these positions can be combined with a different refinement strategy. In the first position, an MC strategy (i.e., setting free loading values initially specified to be zero in the target) may be more advisable, whereas in the second position an MS strategy (i.e., setting loading values initially specified to be free values in the target) may be more useful. The CR would be an exploratory option even when a strong target has been defined.

The authors' experience suggests that proposals such as the present one can only be applied in practice when they are implemented in user-friendly and easily available software. For this reason, the procedure proposed here has been implemented in the 10.7 version of the program FACTOR (Ferrando & Lorenzo-Seva, 2017). To compute a RETAM approach with FACTOR, the user has to provide an initial (partially specified) target matrix and determine the refinement model to be applied. In addition, a cross-validation assessment based on split-half random subsamples can be selected.

References

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

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

Ayr, L. K., Yeates, K. O., Taylor, H. G., & Browne, M. (2009). Dimensions of postconcussive symptoms in children with mild traumatic brain injuries. *Journal of the International Neuropsychological Society*, *15*, 19–30. https://doi.org/10.1017/S1355617708090188

Browne, M. W. (1972a). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 207–212. https://doi.org/10.1111/j.2044-8317.1972.tb00492.x

Browne, M. W. (1972b). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 115–120. https://doi.org/10.1111/j.2044-8317.1972.tb00482.x

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

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*, 230–258. https://doi.org/10.1177/0049124192021002005

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, *31*, 33–42. https://doi.org/10.1007/BF02289455

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cureton, E. E., & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika*, *40*, 183–195. https://doi.org/10.1007/BF02291565

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.

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

Ferrando, P. J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor-analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling*, *16*, 364–381. https://doi.org/10.1080/10705510902751374

Gruvaeus, G. T. (1970). A general approach to Procrustes pattern rotation. *Psychometrika*, *35*, 493–505. https://doi.org/10.1007/BF02291822

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Mathematical and Statistical Psychology*, *17*, 65–70. https://doi.org/10.1111/j.2044-8317.1964.tb00244.x

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

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200. https://doi.org/10.1007/BF02289233

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

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, *34*, 347–356. https://doi.org/10.1207/S15327906MBR3403_3

Lorenzo-Seva, U. (2003). A factor simplicity index. *Psychometrika*, *68*, 49–60. https://doi.org/10.1007/bf02296652

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*, 57–64. https://doi.org/10.1027/1614-2241.2.2.57

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, *109*, 502–511. https://doi.org/10.1037/0033-2909.109.3.502

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*, 611–637. https://doi.org/10.1207/s15327906mbr3604_06

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84–99. https://doi.org/10.1037/1082-989x.4.1.84

McDonald, R. P. (2005). Semiconfirmatory factor analysis: The example of anxiety and depression. *Structural Equation Modeling*, *12*, 163–172. https://doi.org/10.1207/s15328007sem1201_9

Moore, T. M. (2013). *Iteration of target matrices in exploratory factor analysis*. Available from ProQuest Dissertations & Theses Global.

Moore, T. M., Reise, S. P., Depaoli, S., & Haviland, M. G. (2015). Iteration of partially specified target matrices: Applications in exploratory and Bayesian confirmatory factor analysis. *Multivariate Behavioral Research*, *50*, 149–161. https://doi.org/10.1080/00273171.2014.973990

Mosier, C. I. (1939). Determining a simple structure when loadings for certain tests are known. *Psychometrika*, *4*, 149–162. https://doi.org/10.1007/BF02288493

Myers, N. D., Ahn, S., & Jin, Y. (2013). Rotation to a partially specified target matrix in exploratory factor analysis: How many targets? *Structural Equation Modeling*, *20*, 131–147. https://doi.org/10.1080/10705511.2013.742399

Myers, N. D., Ahn, S., Lu, M., Celimli, S., & Zopluoglu, C. (2017). Reordering and reflecting factors for simulation studies with exploratory factor analysis. *Structural Equation Modeling*, *24*, 112–128.

Myers, N. D., Jin, Y., Ahn, S., Celimli, S., & Zopluoglu, C. (2015). Rotation to a partially specified target matrix in exploratory factor analysis in practice. *Behavior Research Methods*, *47*, 494–505. https://doi.org/10.3758/s13428-014-0486-7

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.

Tucker, L. R. (1944). A semi-analytical method of factorial rotation to simple structure. *Psychometrika*, *9*, 777–792. https://doi.org/10.1007/BF02288713

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

Vigil-Colet, A., Morales-Vives, F., Camps, E., Tous, J., & Lorenzo-Seva, U. (2013). Desarrollo y validacion de las Escalas de Evaluacion Global de la Personalidad (OPERAS). *Psicothema*, *25*, 100–107.

Woods, S. A., & Anderson, N. R. (2016). Toward a periodic table of personality: Mapping personality scales between the five-factor model and the circumplex model. *Journal of Applied Psychology*, *101*, 582–604. https://doi.org/10.1037/apl0000062