Psychological Test Toolbox: A New Tool to Compute Factor Analysis Controlling Response Bias

David Navarro-Gonzalez
Universitat Rovira i Virgili

Andreu Vigil-Colet
Universitat Rovira i Virgili

Pere Joan Ferrando
Universitat Rovira i Virgili

Urbano Lorenzo-Seva
Universitat Rovira i Virgili

Abstract

The effects of response bias in psychological tests have been investigated for years, the two most common being social desirability (SD) and acquiescence (AC). However, the traditional methods for controlling or eliminating the impact of those biases in participants' scores have several limitations. Some factor analysis-based methods can overcome some of these limitations, such as the procedure proposed by Ferrando, Lorenzo-Seva, and Chico (2009). Nevertheless, this method involves programming skills that are not common among applied researchers or clinicians. Consequently, we have developed a stand-alone, user-friendly application that provides an easy way of using the aforementioned method to perform a factor analysis which controls for the effect of AC and SD. The program has been developed in the MATLAB environment and its distribution is entirely free.

Keywords: response bias, social desirability, acquiescence, exploratory factor analysis, MATLAB.

1. Introduction

The present paper concerns the exploratory factor analysis of psychological tests. In a typical psychological test, the person being tested responds to a number of items by stating how well each item describes him/her. The responses to these kind of self-reports are susceptible to response bias, which is a systematic tendency to answer the items on some other basis than the specific item content (Paulhus 1991). The two best known response biases in questionnaires are acquiescence (AC) and social desirability (SD). Acquiescence can be defined as the tendency of responders to agree with a statement regardless of its content (Paulhus and Vazire

2005), while social desirability can be defined as people's tendency to present themselves in a generally favorable fashion (Holden 2010).

A review of the literature on response biases indicates that AC and SD can impact both the individual scores and the factorial structure of typical response measures such as personality traits (Danner, Aichholzer, and Rammstedt 2015; Navarro-Gonzalez, Lorenzo-Seva, and Vigil-Colet 2016; Rammstedt and Kemper 2011; Rammstedt and Farmer 2013; Soto, John, Gosling, and Potter 2008) or aggression measures (Becker 2007; Harris 1997; Vigil-Colet, Ruiz-Pamies, Anguiano-Carrasco, and Lorenzo-Seva 2012), among others. The distortions generated by these response biases can have a negative impact on personnel selection, individual assessment, research studies and clinical evaluation (Goffin and Christiansen 2003; Salgado 2005; Viswesvaran, Ones, and Hough 2001).

Given these findings, test developers often use some type of procedure to control or minimize the effect of AC and SD when designing questionnaires. However, most of these procedures are purely descriptive and have some shortcomings due to the ad hoc approaches inherent in those methods. In recent years, more sophisticated model-based procedures have been proposed.

Regarding AC, several authors have proposed procedures based on a factor analysis (FA) model that uses fully balanced sets of items, where half of the items measure in one direction of the trait and the other half measure in the opposite direction. Some procedures are based on restricted factor analysis (Billiet and McClendon 2000; Mirowsky and Ross 1991) and others are based on unrestricted factor analysis Ferrando, Lorenzo-Seva, and Chico 2003; Lorenzo-Seva and Rodríguez-Fornells 2006; Berge 1999). However, in applied research it is usual to find scales that are only partially balanced, which makes the application of the aforementioned procedures difficult. To overcome this limitation, Lorenzo-Seva and Ferrando (2009) proposed a new method for controlling AC in partially balanced multidimensional scales.

In the case of SD, there have traditionally been two different approaches for dealing with its bias effects. Both approaches are based on administering an SD scale together with the content scales. The first method consists of using the SD scale to remove individuals with high scores in SD. This procedure has some limitations. First, removing participants with high scores in SD does not guarantee that the scores of the other participants are free of SD. Second, if the content that is being measured is related to SD, then individuals with high content scores may also be removed.

The second traditional method is known as "partialling", which is based on using the SD scale to partial out the SD effects on the content scale by regressing the SD scores onto the trait scales of interest and computing a residual score. This approach also has some limitations. First, it may remove meaningful variance from the relevant trait. Second, the procedure assumes that all items are parallel measures of the trait of interest, which is almost never true.

These limitations may be overcome by using methods based upon factor analysis. Some FA-based procedures for identifying an SD factor are those proposed by Paulhus (1981) or by Neill and Jackson (1970). In particular, in the procedure suggested by Neill and Jackson (1970) SD is identified by using an SD scale as a marker variable. Ferrando (2005) developed a restricted FA model by expanding Jackson's idea to assess simultaneously the effects of AC and SD, thus allowing these biases to be modeled as additional factors that can be distinguished from

the content factors (Ferrando *et al.* 2009). This procedure presents certain advantages. First, it removes the effect of both response biases from the factor structure, allowing the item structure to be analyzed once the distortion generated by SD and AC is removed. Second, it provides the estimated factor scores of the participants, which is a more precise and unbiased estimation of how the individuals stand with regard to the trait that is to be measured, and this can be very useful in individual assessment.

The main practical application of the procedure is its use in the development of new questionnaires, but it can also be applied a posteriori if the analyzed questionnaire meets the characteristics described in Section 2.

The procedure has been considered of interest in applied research and it has been successfully used to develop different questionnaires (Cupani and Lorenzo-Seva 2016; Mas-Herrero, Marco-Pallares, Lorenzo-Seva, Zatorre, and Rodriguez-Fornells 2013; Morales-Vives, Camps, and Lorenzo-Seva 2013; Ruiz-Pamies, Lorenzo-Seva, Morales-Vives, Cosi, and Vigil-Colet 2014; Vigil-Colet, Morales-Vives, Camps, Tous, and Lorenzo-Seva 2013). However, only researchers with advanced knowledges of psychometrics and programming can perform such an analysis, and this may hinder the wider application of the method.

Taking this into account we have developed a stand-alone application called **Pychological Test Toolbox**, which is a user-friendly application that enables the implementation of the procedure described in Ferrando *et al.* 2009. It provides an easy way of performing factor analysis by controlling the effect of AC and SD, thus providing bias-free individual response scores. The program has been developed in the MATLAB environment (The MathWorks Inc. 2017).

R (R Core Team 2019) is a popular statistical software package among researchers because it is open-source and can be used for free. We have implemented our software in MATLAB following the same philosophy as in R: any user can download and use it as a freeware software. We decided to use MATLAB because it is simpler to produce user-friendly software. The **Psychological Test Toolbox** is not compatible with Octave (Eaton, Bateman, Hauberg, and Wehbring 2019) because Octave lacks some of the core functions required for the GUI's tab system and some other minor functions.

2. Characteristics of psychological tests to control SD and AC

The procedure proposed by Ferrando *et al.* (2009) cannot be applied to any typical response measure. In order to control SD, a number of items related to SD must be included in the psychological test. These items are known as SD markers. The greater the number of markers, the better the procedure is expected to work. However, in applied research the procedure has been successfully applied with as few as four SD markers. The psychological test (or questionnaire) must therefore be composed of (a) a short number of SD markers (at least four), and (b) the setoff items related to the psychological latent variables that the psychological test aims to assess. The procedure allows more than one latent variable to be assessed and they can be correlated.

In order to deal with AC, the procedure assumes that it should be possible to identify acquiescence as a common style factor behind a set of content items that are semantically balanced (Mirowsky and Ross 1991). In a perfectly balanced scale, with respect to a psychological trait, half of the items are worded in one direction and the other half in the other. However, few questionnaires are designed so that exactly half of the items are worded in this way:

Most of the psychological tests are only partially balanced. Fortunately, the procedure by Lorenzo-Seva and Ferrando (2009) helps to handle partially balanced scales (i.e., where only a few items in the scale are worded in the opposite direction). In partially balanced scales, two subsets of content items must be identified: (a) a balanced subset (i.e., a subset of items where half of the items are worded in one direction and the other half in the other); and (b) an unbalanced subset (i.e., a subset of items where all the items are worded in the same direction). It must be noted that the procedure finally removes the variance caused by AC from all the items in the questionnaire (i.e., from the balanced subset of items, but also the unbalanced subsets of items).

An example of a psychological test that includes SD markers and partially balanced content items is OPERAS (Vigil-Colet _et al._ 2013). This test includes 4 SD markers, and 35 content items of which 18 are reversed. This psychological test assesses the individuals' scores for 5 latent variables. It must be noted that a psychological test that does not include SD markers but which does have (partially) balanced content items can only control for AC; whereas a psychological test that includes a number of SD markers, but with all the content items worded in the same direction, can only control SD. Finally, the procedure proposed by Ferrando _et al._ (2009) is based on two strong assumptions: (a) AC and SD measures are assumed to be uncorrelated from the content factors and from each other; and (b) AC is assumed not to operate in pure SD markers. As a consequence, SD and AC can be controlled in consecutive and separate steps. In Section 3.1 we present how SD can be controlled using the SD markers, and in Sections 3.2 and 3.3 we describe how AC can be controlled in partially balanced scales. In Section 4 we discuss some of the existent related software packages. In Section 5 we present our stand-alone package for computing the procedure.

3. Model overview

3.1. Controlling social desirability

Let us consider a questionnaire administrated to n individuals and composed of m content items. The m items are a set of items expected to be related to r latent content variables ($r < m$). The questionnaire is partially balanced: A subset of k items is worded in one direction of the trait, and a subset of l items is worded in the opposite direction, where $k + l = m$. Additionally, a set of h SD markers are administrated together with the content items. The $\mathbf{X}$ matrix containing scores of the n individuals (i.e., the responses of individuals to the test) can be partitioned as

$$\mathbf{X} = [\mathbf{X}^{sd}|\mathbf{X}^c], \tag{1}$$

where $\mathbf{X}^{sd}$ contains the scores in the SD markers and $\mathbf{X}^c$ contains scores related to all the m content items. $\mathbf{X}^c$ can be partitioned as

$$\mathbf{X}^c = [\mathbf{X}^b|\mathbf{X}^u], \tag{2}$$

where $\mathbf{X}^b$ contains scores related to an even set of k balanced items, and where $\mathbf{X}^u$ contains scores related to a set of l unbalanced items. The correlation between all the items included in $\mathbf{X}$ will be contained in $\mathbf{R}$. Also, $\mathbf{R}^c$ contains the correlation between $\mathbf{X}^c$ items and $\mathbf{R}^{sd}$ contains the correlation between $\mathbf{X}^{sd}$ items

The structural model assumes that each content item is a factorially complex measure, determined by: (a) the SD factor θ^{sd}, (b) an AC factor θ^a, and (c) the r content factors θ^c

$$X_{ij} = \delta_j \theta_i^{sd} + \alpha_j \theta_i^a + \beta_{j1} \theta_{i1}^c + \beta_{j2} \theta_{i2}^c + \ldots + \beta_{jr} \theta_{ir}^c + \varepsilon_{ij} \tag{3}$$

for $i = 1 \ldots n$ and $j = 1 \ldots m$, where δ is the SD factor loading, α is the AC factor loading, βs are the factor loadings for r content factors and the εs are the residuals, with zero means and uncorrelated with the factors or one another. As mentioned above, the r factors are assumed to be uncorrelated with the response bias factors. Also, the SD factor and AC factor are also expected to be uncorrelated with each other. However, the r content factors can be correlated among themselves.

To simplify the model, let us suppose that all content items in the questionnaire are measuring a one-dimensional trait θ^c, thus leading to a model such as

$$X_{ij} = \delta_j \theta_i^{sd} + \alpha_j \theta_i^a + \beta_j \theta_i^c + \varepsilon_{ij}. \tag{4}$$

Consider now the additional set of h items designed to be pure measures of SD, which are administrated together with the content items. Their function is to provide factorially simple measures of SD, and the structural model for these items reduces to:

$$X_{ih} = \delta_h \theta_i^{sd} + \varepsilon_{ih}. \tag{5}$$

The h SD markers allow the loading of the content items on the SD factor to be estimated using the instrumental variables (IV) technique. This technique was developed in the context of factor analysis by Hägglund (1982). First, one of the SD markers is taken as a proxy for θ^{sd} and the remaining $h-1$ markers are taken as instrumental variables. Without loss of generality we can take the first marker as proxy. From correlation matrix $\mathbf{R}$, two vectors $\mathbf{r}_h$ and $\mathbf{r}_j$ can be defined. $\mathbf{r}_h$ is a column vector of order $(h-1) \times 1$ that contains the covariance between the proxy and the other $h-1$ markers. $\mathbf{r}_j$ is a column vector of order $(h-1) \times 1$ that contains the covariance between the content item j and the other $h-1$ markers. Then the loading of the m content items on the SD factor can be estimated as,

$$\hat{\delta}_j = \mathbf{r}_j^\top \mathbf{r}_h (\mathbf{r}_h^\top \mathbf{r}_h)^{-1} \delta_1. \tag{6}$$

where $\hat{\delta}_j$ is the loading of the content item j, and δ_1 is the factor loading of the proxy variable. The value of δ_1 can be computed from the correlation matrix of the h SD markers, or directly defined from a previous study.

This is how the loadings of the SD factor for the m content items can be estimated. The loadings for the $h-1$ SD markers are estimated in the same way, and the loading for the first marker (proxy) can be estimated simply by choosing another pivot variable. Once the complete vector of m loading estimates $\boldsymbol{\delta}$ have been obtained, the reproduced correlation matrix is computed as $\boldsymbol{\delta}\boldsymbol{\delta}^\top$.

The first residual matrix $\mathbf{S}^c$, which is free of SD impact, is obtained by substracting the reproduced matrix from the initial correlation matrix between the content items $\mathbf{R}^c$, defined as

$$\mathbf{S}^c = \mathbf{R}^c - \boldsymbol{\delta}\boldsymbol{\delta}^\top. \tag{7}$$

3.2. Controlling acquiescence: Method for fully balanced scales

This first residual matrix obtained after subtracting SD variance is used as the input in the second stage for estimating the loadings on the AC factor. As the influence of the SD factor has been partialled out, the structural model looks like this:

$$X_{ij} = \alpha_j \theta_i^a + \beta_j \theta_i^c + \varepsilon_{ij}. \tag{8}$$

If $\mathbf{S}^c$ is the first residual matrix obtained after substracting SD variance of the order $m \times m$, $\mathbf{S}^b$ is the residual matrix between a set of balanced items. Then

$$\mathbf{a} = \mathbf{S}^b \mathbf{1}(\mathbf{1}^\top \mathbf{S}^b \mathbf{1})^{-1/2}, \tag{9}$$

where $\mathbf{a}$ is the vector of correlations between the variables and their mean. Values of $\mathbf{a}$ show how much each variable is impacted by AC. A factor loading matrix $\mathbf{B}_b$ of the order of $m \times (r+1)$ can be obtained by

$$\mathbf{S}^b = \mathbf{B}_b \mathbf{B}_b^\top + \mathbf{M}_b \mathbf{M}_b^\top + \mathbf{\Psi}_b^2, \tag{10}$$

where $\mathbf{M}_b$ holds the loadings on those common factors that are discarded in the rank-$(r+1)$ solution and $\mathbf{\Psi}_b$ is a diagonal matrix containing the unique factor standard deviations. Let the rotation matrix $\mathbf{W}$ be an orthonormal matrix of order $(r+1) \times (r+1)$. $\mathbf{W}$ must maximize the congruence between one column of the product $\mathbf{M}_b \mathbf{W}$ and vector $\mathbf{a}$, so it is determined by the method of Korth and Tucker (1976). Let $\mathbf{d}$ and $\mathbf{w}$ be vectors defined as

$$\mathbf{d} = (\mathbf{B}_b^\top \mathbf{B}_b)^{-1} \mathbf{B}_b^\top \mathbf{a} \tag{11}$$

and

$$\mathbf{w} = \mathbf{d}(\mathbf{d}^\top \mathbf{d})^{-1/2}. \tag{12}$$

Given the eigendecomposition of the matrix

$$\mathbf{I} - \mathbf{w}\mathbf{w}^\top = \mathbf{W}\Delta\mathbf{W}^\top, \tag{13}$$

where $\mathbf{I}$ is an identity matrix and Δ is a diagonal matrix with elements $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_{r+1} = 0$, the product

$$\mathbf{B}_b \mathbf{W} = [\boldsymbol{\beta}|\boldsymbol{\alpha}] \tag{14}$$

leads to a matrix whose last column $\boldsymbol{\alpha}$ contains the loading values of balanced items on the acquiescence factor, and $\boldsymbol{\beta}$ is a $k \times r$ matrix that can be rotated to show factor simplicity by any orthogonal or oblique rotation method. Note that $\boldsymbol{\beta}$ is a factor loading matrix that is free of variance caused by AC responding.

3.3. Controlling acquiescence: Method for partially balanced scales

A factor loading matrix $\mathbf{L}$ of the order of $m \times (r+1)$ can be obtained by

$$\mathbf{S}^c = \mathbf{L}\mathbf{L}^\top + \mathbf{M}\mathbf{M}^\top + \mathbf{\Psi}^2, \tag{15}$$

where $\mathbf{S}^c$ is the covariance matrix obtained after substracting SD variance, $\mathbf{M}$ holds the loadings on the common factors that are discarded in the rank-$(r+1)$ solution and $\mathbf{\Psi}$ is a diagonal matrix containing that unique factor standard deviations. $\mathbf{L}$ can be portioned as

$$\mathbf{L} = \left[\frac{\mathbf{L}_b}{\mathbf{L}_u} \right], \tag{16}$$

where $\mathbf{L}_b$ contains the loading values related to the even set of balanced items and $\mathbf{L}_u$ contains the loading values related to the set of unbalanced items. Let the rotation matrix $\mathbf{U}$ be an orthonormal matrix of order $(r+1) \times (r+1)$. $\mathbf{U}$ must maximize the congruence between one column of the product $\mathbf{L}_b\mathbf{U}$ and vector $\mathbf{a}$, so it is determined by the method of Korth and Tucker (1976). Finally, $\mathbf{U}$ is used to rotate not only $\mathbf{L}_b$ but also the overall matrix $\mathbf{L}$, so that the product

$$\mathbf{L}\mathbf{U} = [\boldsymbol{\beta}|\boldsymbol{\alpha}] \tag{17}$$

leads to a matrix whose last column $\boldsymbol{\alpha}$ contains the loading values of balanced and unbalanced items on the acquiescence factor and $\boldsymbol{\beta}$ is an $m \times r$ matrix that can be rotated to show factor simplicity by any orthogonal or oblique rotation method. If $\mathbf{T}_r$ is an $r \times r$ rotation matrix, the rotated loading matrix related to the content factors is obtained by

$$\mathbf{P}_r = \boldsymbol{\beta}\mathbf{T}_r, \tag{18}$$

while the correlation matrix between factor scores is obtained by

$$\boldsymbol{\Phi}_r = \mathbf{T}_r^{-1}(\mathbf{T}_r^{-1})^{\top}. \tag{19}$$

The hull procedure is summarized in Figure 1.

4. Available software packages

Factor analysis is implemented in most software packages. As a stand-alone package, among the most widely used freeware packages is **FACTOR** (Lorenzo-Seva and Ferrando 2013), which implements several methods for computing factor analysis including some of the most recent methodological developments. The most common R distribution package for computing factor analysis is probably the **psych** package (Revelle 2019), which also contains several configuration options, and is up to date in methodological developments. Both options are really good tools for computing FA, and are clearly more configurable than the **Psychological Test Toolbox** in terms of the number of procedures available for the user to choose. However, none of them are able to control response biases in their procedures, which is the main reason that we created our tool.

Regarding the response bias function, there are certain other factor analysis procedures for assessing the impact of acquiescence or social desirability. However, to the best of our knowledge, none of those methods are available for distribution via an R package or any other software. The only way some of the aforementioned methods can be used is by manually calculating all the computing steps with the equations provided in the articles.

The only similar tools for controlling response bias are correctors in specific instruments, which are designed solely to provide the participant's factor scores for the specific version of a given questionnaire.

5. Psychological Test Toolbox

5.1. Overall description

As mentioned previously, **Psychological Test Toolbox** is a program designed for performing factor analysis while controlling the effect of both AC and SD or only one of these biases. The
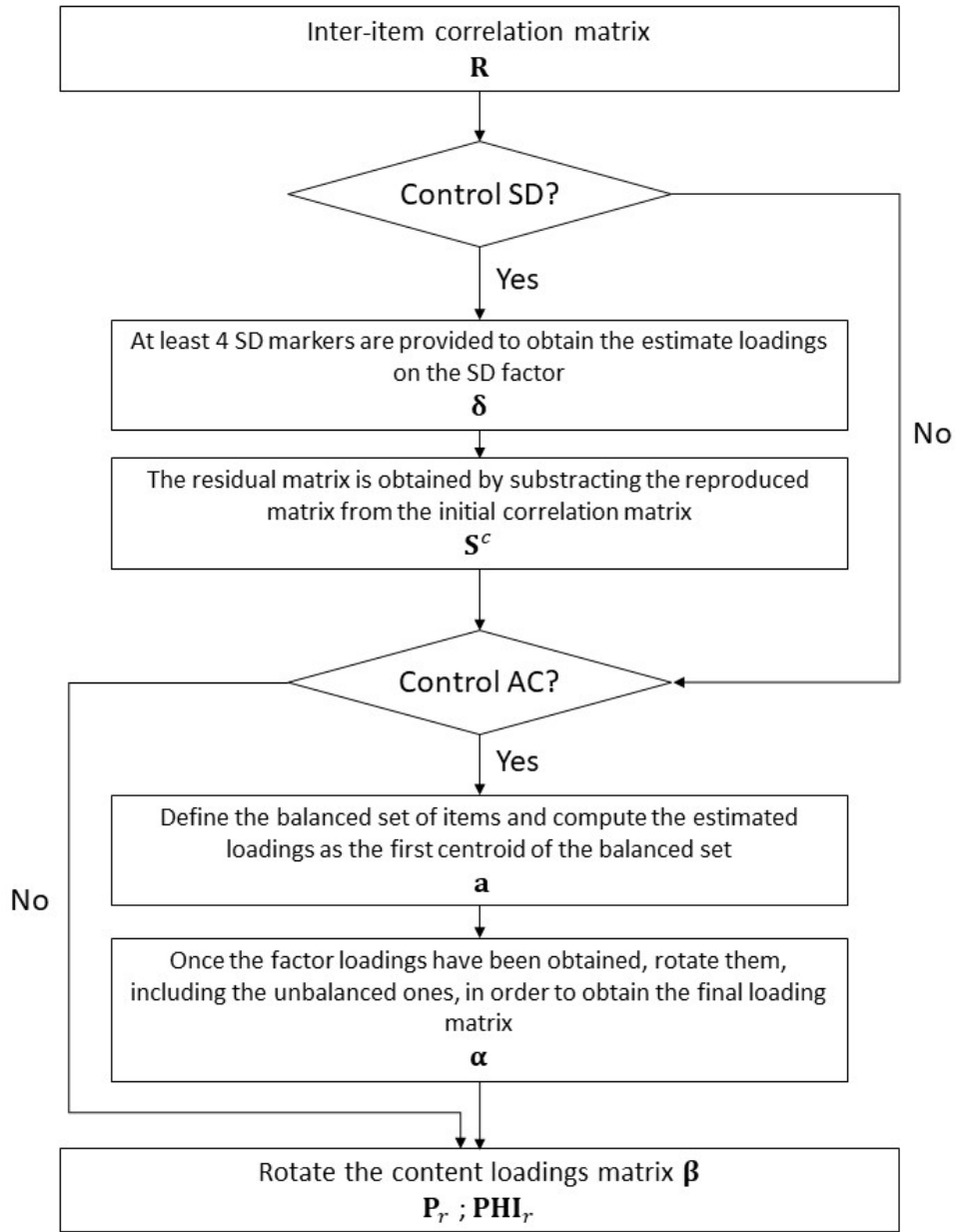
Figure 1: Procedure for controlling social desirability bias, acquiescence bias or both.

program was developed in the MATLAB environment, and it is released in two formats: as stand-alone application (only for Windows-based computers) and also as a MATLAB toolbox, which can be executed by MATLAB users on any operating system which supports MATLAB.

The program is free, it only requires the installation of MATLAB Compiler Runtime (MCR) for R2019b, which is available for free from the MathWorks® website.

5.2. Procedures implemented

It is important to mention that this project implements more than one hundred functions,
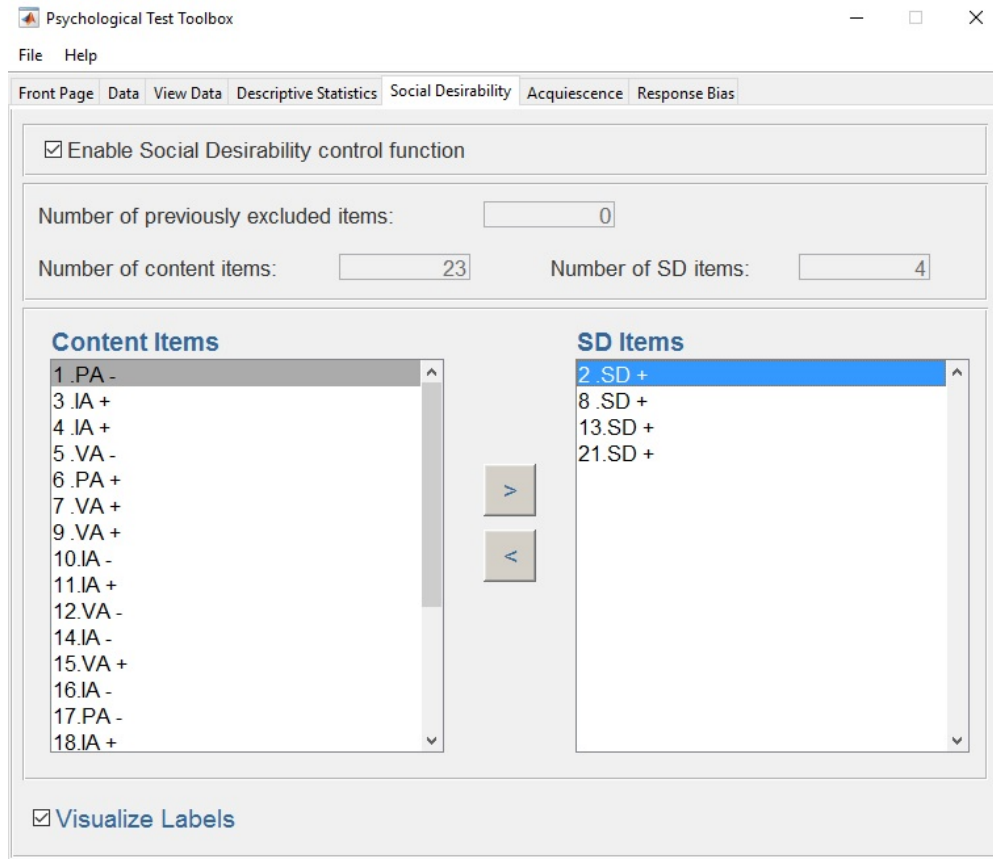
Figure 2: Determining the social desirability markers.

including the primary and the secondary ones (invoked by the primaries). So it is not practical to list them all in this article. However, we make an effort to comment each one in the code, especially the primaries, to describe their usability. Also, in the principal script (`PsychologicalTestToolbox.m`), all the objects and the functions embedded have a comment line to guide the MATLAB user during the calculation process.

To obtain a summary of the functions used in the program, the MATLAB user can use the following command lines in the MATLAB prompt:

```
addpath PsychologicalTestToolbox/lib/
help lib
```

Regarding the authorship of the functions used in the program, the vast majority of them are entirely written by members of our research team. The only exception are some internal computing functions present in the polychoric matrix calculation, which were originally created by Beasley and Springer (1977), Brown (1977) and Donnelly (1973). Also, if the code is based on a method proposed by a certain author, this is mentioned in the code itself or in the reference section if the contribution to the calculation is significant

The program allows to compute factor analysis using different kinds of dispersion matrices, including covariance matrices, Pearson correlation matrices and tetrachoric/polychoric cor-

relation matrices. The suitability of the dispersion matrix is assessed by three tests: the determinant of the matrix, Barlett's test, and the Kaiser-Meyer-Olkin index.

The number of factors to be retained has to be specified, and the optimal implementation of parallel analysis (PA, Timmerman and Lorenzo-Seva 2011) can be computed to assess which number of factors is suitable. The eigenvalues of the dispersion matrices and Cattell's scree test are also generated.

For factor analysis, the program uses two procedures: unweighted least squares (ULS) and minimum rank factor analysis (MRFA, Berge and Kiers 1991). For assessing the model's goodness of fit, the program provides the goodness of fit index (GFI), the root mean square of residuals (RMSR), and descriptive statistics of the distribution of the residuals.

The program includes the following rotation methods: Varimax (Kaiser 1958) and Promin (Lorenzo-Seva 1999), which is a special case of simplimax (Kiers 1994), for assessing orthogonal and oblique solutions, respectively. For assessing semi-specified target rotation it provides the methods developed by Browne (1972a) and Browne (1972b).

After the rotation phase, the **Psychological Test Toolbox** provides Bentler's simplicity index (Bentler 1977) and Lorenzo's-Seva simplicity index (Lorenzo-Seva 2003) to assess the level of simplicity attained in the rotated solution. Also, if the target matrix is provided, the congruence indices between the rotated solution and the expected solution are given, thus providing the congruence for each item and for each factor as well as the overall congruence.

Factor scores are computed by an improved implementation of Bayes EAP (expected a posteriori) estimation described in Ferrando and Lorenzo-Seva (2016), which also provides the standard error of prediction for all responders.

Missing values are processed using the multiple imputation of missing values described in Lorenzo-Seva and Ginkel (2016).

5.3. Design of the graphical user interface

The design of the graphical user interface (GUI) was one of the most important phases in the development process because one of the main objectives of the **Psychological Test Toolbox** was to create a very accessible program. We tried to develop a simple GUI using the tools that are provided by the MATLAB language and divided the hull application into 7 tabs which are organized according to the logical order of the analyzing process. The name and description of each tab are as follows:

Front Page: This is the first tab displayed and indicates the name of the program, the authors and the current version of the program.

Data: The user uses this tab to import the data that will be used in the analysis. Once it is imported, the user can exclude certain items.

View Data: This tab displays the imported data, that could be useful for doing some checking without having to open the file externally.

Descriptive Statistics: The user can make certain changes to the configuration of the analysis such as changing the dispersion matrix that will be used or setting up parallel analysis. Also, the descriptive statistics section can be computed and displayed in this

tab, which could be useful in certain cases if the user is not interested in computing a factor analysis at that particular moment.

Social Desirability: In this tab the user can enable the SD control function. If enabled, the application requires the user to select at least four items as SD markers.

Acquiescence: In this tab the user can enable the AC control function. If enabled, the application provides the user with the option of excluding certain items from the balanced core of items. As explained in Section 2, the questionnaire has to be at least partially balanced in order to control AC.

Response bias: This is the final tab, which includes a complete report of the data to be analyzed including the items excluded from the analysis and those selected as SD markers. In this tab the user has to specify the number of factors to be retained; at least 3 items are required per factor. Also, the user can switch between the rotation methods available, require the participant's factor scores to be computed and require all possible bias combinations to be computed. Once the analysis is complete, the output is displayed in the embedded sub-window and can be saved for external viewing.

Finally, the program has a "Help" menu and a "File" menu, the latter featuring certain options such as importing data, saving matrices generated during the analysis or exiting. One of the functions available in the "File" menu is saving the current configuration of the program, including the data imported and the output generated (if any), and allowing the user to close the program and resume the analysis at a later time point by clicking on "Save analysis" and "Open analysis". This could be a useful tool for replicating certain results using the same exact configuration, or for doing a complex analysis at two different time points.

All the GUI objects, figures and graphics are generated by code, which gave us more flexibility to handle and structure them. The main figure where the application is embedded cannot be resized to prevent distortions of the objects from being viewed. The figures containing plots and the output section are fully resizable.

5.4. Input and output

To run the stand-alone program, **Psychological Test Toolbox** must be executed on a Windows operating system. To run the user-friendly interface in MATLAB as a toolbox, the following command line must be executed on the MATLAB prompt:

```
PsychologicalTestToolbox
```

Once the main window is in execution, the program requires some input data to work with, which can be a raw data file or a dispersion matrix. The program can import files in different formats (`.dat`, `.txt`, `.xls`, `.xlsx`), and is able to identify variable labels in the header of the file. There are some optional input files, such as a text file containing the variable labels (in `.txt`) and a file containing the semi-specified rotation target matrix.

If the data contains missing values, the user has to assign a unique value to these (for example: 999), specify the option "The data file contains missing values" and define the previously determined value.

An extensive output is provided, depending on the selected options. The output is divided into two parts. The first output section contains the descriptive statistics of the items, including:
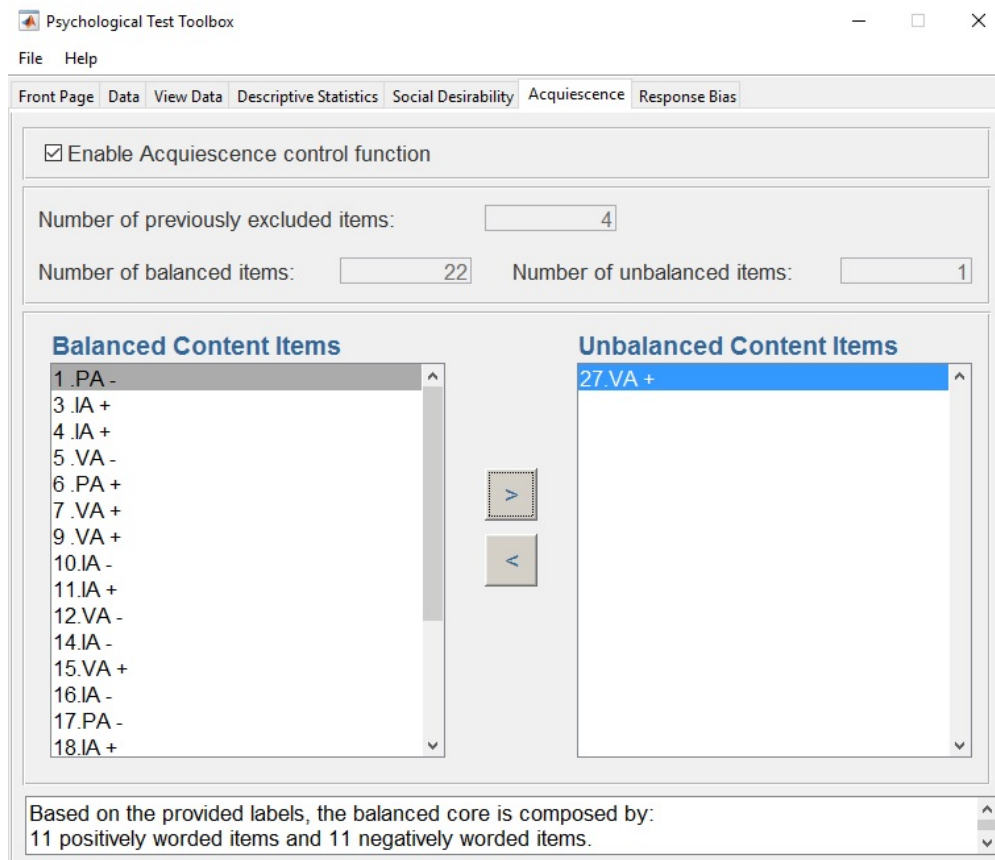
Figure 3: Selecting the balanced core of items.

- a summary of the analysis;

- univariate item descriptive statistics;

- dispersion matrix;

- parallel analysis output (if requested);

- indices of adequacy of the dispersion matrix;

- scree test;

- descriptive statistics related to missing data (if applicable);

- references for this section.

And the second output section presents the factor analysis output, including:

- goodness of fit index;

- target loading matrix (if provided);
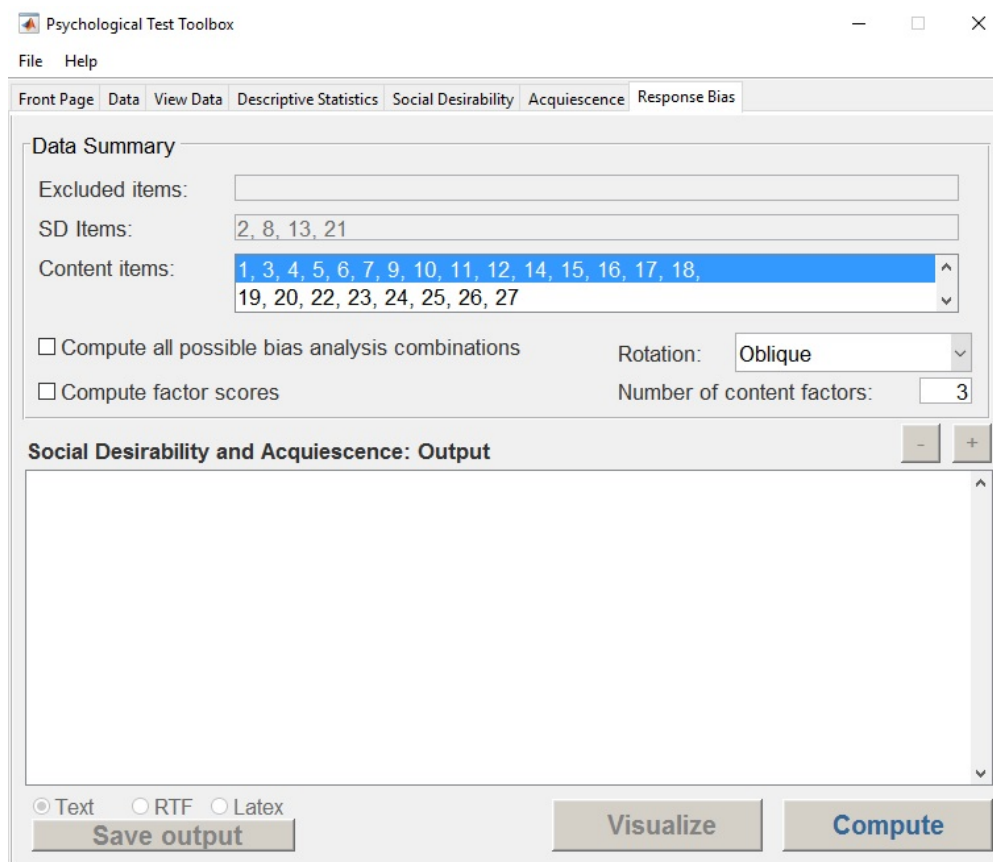
- rotated loading matrix;

Figure 4: The configuration screen which displays the list of content items and SD markers.

- correlation between content factors;

- indices of factor simplicity;

- congruence indices between the rotated loading matrix and target matrix (if a Procrustean rotation is selected);

- distribution of residuals;

- EAP scores of the participants and the reliability of these scores (if required);

- references for this section.

Note that the factor analysis output will be generated for the desired option (controlling for only one response bias, or both or neither) but there is an option for computing and displaying all the possible bias analysis combinations.

The output can be saved in three different formats: in plain text (`.txt`), in Rich Text Format (`.rtf`) which is fully compatible with Microsoft Word and presents all the tables in a proper format, and also in LaTeX (`.tex`) format which generates a complete report that presents all the information in a clean manner.

Item	Real-data % of variance	Mean of random % of variance	95 percentile of random % of variance
1	29.09**	7.33	7.84
2	8.78**	6.89	7.31
3	8.25**	6.55	6.91
4	5.84	6.25	6.61
5	5.38	5.96	6.27
6	4.00	5.70	6.02
7	3.59	5.45	5.74
8	3.40	5.20	5.45
9	3.24	4.96	5.19
10	2.81	4.72	4.96
11	2.75	4.48	4.67
12	2.67	4.24	4.44
13	2.48	4.00	4.23
14	2.45	3.76	3.98
15	2.26	3.50	3.71
16	2.15	3.26	3.48
17	1.97	3.00	3.25
18	1.85	2.75	3.00
19	1.61	2.49	2.75
20	1.57	2.22	2.51
21	1.43	1.94	2.22
22	1.04	1.67	1.97
23	0.66	1.38	1.69
24	0.36	1.08	1.42
25	0.21	0.76	1.07
26	0.15	0.45	0.70
27	0.00	0.00	0.00

Table 1:   Parallel analysis output. ** advised number of factors

6. An illustrative example

In this example we are going to use the indirect and direct aggression questionnaire (I-DAQ, Ruiz-Pamies *et al.* 2014), which was one of the first questionnaires developed using this procedure to control for response biases. This questionnaire was administered to a sample of 1479 respondents (536 men and 943 women) with age ranging from 18 to 96 years.

The questionnaire measures 3 aggression dimensions: physical aggression (PA), verbal aggression (VA) and indirect aggression (IA). The questionnaire consists of 27 Likert items, i.e., 23 items measuring the 3 content dimensions and 4 SD markers for applying the procedure described in this article. For clarity, we used labels indicating which dimensions were being measured and their direction; the positively keyed items are labeled "+" and the negatively worded items are labeled "−". The selection of the SD markers in the program is presented in Figure 2.

The content items are only partially balanced and consist of 12 positively worded items and

Item	SD	AC	Physical	Indirect	Verbal
2.SD +	0.6468	0.0000	0.0000	0.0000	0.0000
8.SD +	0.4668	0.0000	0.0000	0.0000	0.0000
13.SD +	0.7851	0.0000	0.0000	0.0000	0.0000
21.SD +	0.7706	0.0000	0.0000	0.0000	0.0000
1.PA −	−0.1806	0.1098	**−0.3863**	−0.1306	−0.2143
6.PA +	0.1495	0.3287	**0.5937**	−0.0771	−0.0380
17.PA −	−0.1381	0.2308	**−0.7347**	0.0121	0.0292
19.PA −	−0.1685	0.2648	**−0.7004**	0.0589	0.0007
20.PA +	0.2985	0.2528	**0.6365**	0.0720	−0.1066
25.PA +	0.2347	0.2109	**0.5231**	0.0566	0.1719
3.IA +	0.4166	0.3004	0.0328	**0.4692**	−0.0080
4.IA +	0.3528	0.2982	0.0875	**0.3860**	−0.0072
10.IA −	−0.0896	0.2345	0.0306	**−0.4247**	0.0408
11.IA +	0.1982	0.3219	−0.0809	**0.5189**	0.0334
14.IA −	−0.2053	0.1864	0.0010	**−0.4552**	−0.1040
16.IA −	−0.0921	0.1572	0.1003	**−0.6288**	0.0242
18.IA +	0.2984	0.3344	0.0876	**0.6144**	−0.1073
23.IA +	0.2689	0.3341	0.0879	**0.5271**	−0.0164
24.IA −	−0.0486	0.1366	0.0460	**−0.4031**	−0.0742
26.IA −	−0.1850	0.1497	−0.0031	**−0.5082**	−0.0271
5.VA −	−0.1620	0.1060	0.0094	0.0047	**−0.6618**
7.VA +	0.1709	0.1797	−0.0731	0.0851	**0.5408**
9.VA +	0.2881	0.2258	0.0394	0.0786	**0.5314**
12.VA −	−0.1913	0.1410	−0.1039	−0.0147	**−0.5356**
15.VA +	0.2508	0.1170	−0.0654	−0.1663	**0.7512**
22.VA −	−0.1609	0.1016	−0.1278	−0.0748	**−0.2232**
27.VA +	0.4490	0.1538	0.1885	0.0767	**0.2739**

Table 2: Rotated loading matrix.

only 11 negatively worded items. In this example, we will exclude item number 27 from the balanced core (see Figure 3).

Finally, Figure 4 presents the hull configuration, including the full list of content items and the configuration options used in this analysis.

We computed an exploratory factor analysis, controlling both biases based on the polychoric interitem correlation matrix, because polychoric correlation is advised when the univariate distributions of ordinal items are asymmetric or with excess of kurtosis, which is the case. If both indices are lower than one in absolute value, then Pearson correlation is advised. The root mean squares of the residuals (RMSR) was .037. An optimal implementation of parallel analysis was computed and the results are shown in Table 1, showing that the advised number of factors to retain are 3.

Table 2 shows the factor solution obtained by controlling for response bias using the procedure described above. When applying the procedure to control for the effect of SD and AC, the factor structure becomes congruent with the expected solution. All the items have their

salient loading on the expected factor, without any factorially complex item, thus resulting in a simple solution.

Another advantage of using this procedure is the ability to look at the loadings on the biases factors in order to determine which items are more impacted by which bias. For example, we can see that some items have high loadings on the SD factor, such as item 27 (0.449), item 3 (0.417) and item 4 (0.353). Also, there are other items with high loadings on the AC factor such as item 18 (0.334) and item 23 (0.334).

7. Program limitations

The program's options for configuring the analysis have been simplified to make the process easier for applied researchers who may be less familiar with some of the psychometric and statistical concepts. However, this decision has left the program with only a few configuration options, and some advanced users may consider them too limited.

The stand-alone version is only available for Windows-based computers. However, the code version of the program can be executed from any OS if the user has a MATLAB license.

There are some analyses which can take several minutes or even hours. This is not the standard, but in some configurations the computing time can increase substantially, from a few seconds to several minutes. The options that can further increase computation time are:

- selecting the polychoric matrix as the dispersion matrix for the analysis;

- requesting optimal implementation of parallel analysis;

- requesting factor scores for each participant.

If the user only wishes to use one of these options the increase in computing time will be acceptable. However, the computing time begins to increase significantly when some of these options are requested in combination, for example if the user selects the polychoric matrix and requests the factor scores for each participant.

Furthermore, computing time can also be increased by certain sample characteristics such as having a large amount of items or participants or missing data.

8. Software availability

As mentioned previously, the **Psychological Test Toolbox** is a freeware program and can be downloaded from the website of our department:

http://psico.fcep.urv.cat/utilitats/PsychologicalTestToolbox/

In the website, the user will find extensive documentation, including tutorial videos organized in sections depending on which functionalities the user is more interested. We strongly recommend to visit the site to stay up to date regarding the current version of the program, as well as knowing about possible new features introduced.

The library required for executing the stand-alone application can be downloaded from the MathWorks website. This is not necessary if the user has a current license for MATLAB 2019b.

https://www.mathworks.com/products/compiler/matlab-runtime.html

Acknowledgments

References

Beasley JD, Springer SG (1977). "Algorithm AS 111: The Percentage Points of the Normal Distribution." *Journal of the Royal Statistical Society C*, **26**(1), 118–121. `doi:10.2307/2346889`.

Becker G (2007). "The Buss-Perry Aggression Questionnaire: Some Unfinished Business." *Journal of Research in Personality*, **41**(2), 434–452. `doi:10.1016/j.jrp.2006.05.004`.

Bentler PM (1977). "Factor Simplicity Index and Transformations." *Psychometrika*, **42**(2), 277–295. `doi:10.1007/bf02294054`.

Berge JMT (1999). "A Legitimate Case of Component Analysis of Ipsative Measures, and Partialling the Mean as an Alternative to Ipsatization." *Multivariate Behavioral Research*, **34**(1), 89–102. `doi:10.1207/s15327906mbr3401_4`.

Berge JMT, Kiers HAL (1991). "A Numerical Approach to the Approximate and the Exact Minimum Rank of a Covariance Matrix." *Psychometrika*, **56**(2), 309–315. `doi:10.1007/bf02294464`.

Billiet JB, McClendon MJ (2000). "Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items." *Structural Equation Modeling*, **7**(4), 608–628. `doi:10.1207/s15328007sem0704_5`.

Brown MB (1977). "Algorithm AS 116: The Tetrachoric Correlation and Its Asymptotic Standard Error." *Journal of the Royal Statistical Society C*, **26**(3), 343–351. `doi:10.2307/2346985`.

Browne MW (1972a). "Oblique Rotation to a Partially Specified Target." *British Journal of Mathematical and Statistical Psychology*, **25**(2), 207–212. `doi:10.1111/j.2044-8317.1972.tb00492.x`.

Browne MW (1972b). "Orthogonal Rotation to a Partially Specified Target." *British Journal of Mathematical and Statistical Psychology*, **25**(1), 115–120. `doi:10.1111/j.2044-8317.1972.tb00482.x`.

Cupani M, Lorenzo-Seva U (2016). "The Development of an Alternative IPIP Inventory Measuring the Big-Five Factor Markers in an Argentine Sample." *Personality and Individual Differences*, **91**, 40–46. `doi:10.1016/j.paid.2015.11.051`.

Danner D, Aichholzer J, Rammstedt B (2015). "Acquiescence in Personality Questionnaires: Relevance, Domain Specificity, and Stability." *Journal of Research in Personality*, **57**, 119–130. `doi:10.1016/j.jrp.2015.05.004`.

Donnelly T (1973). "Algorithm 462: Bivariate Normal Distribution." *Comunications of the ACM*, **16**(10), 638. `doi:10.1145/362375.362414`.

Eaton JW, Bateman D, Hauberg S, Wehbring R (2019). *GNU Octave Version 5.1.0 Manual: A High-Level Interactive Language for Numerical Computations.* URL `https://www.gnu.org/software/octave/`.

Ferrando PJ (2005). "Factor Analytic Procedures for Assessing Social Desirability in Binary Items." *Multivariate Behavioral Research*, **40**(3), 331–349. `doi:10.1207/s15327906mbr4003_3`.

Ferrando PJ, Lorenzo-Seva U (2016). "A Note on Improving EAP Trait Estimation in Oblique Factor-Analytic and Item Response Theory Models." *Psicológica*, **37**, 235–247.

Ferrando PJ, Lorenzo-Seva U, Chico E (2003). "Unrestricted Factor Analytic Procedures for Assessing Acquiescent Responding in Balanced, Theoretically Unidimensional Personality Scales." *Multivariate Behavioral Research*, **38**(3), 353–374. `doi:10.1207/s15327906mbr3803_04`.

Ferrando PJ, Lorenzo-Seva U, Chico E (2009). "A General Factor-Analytic Procedure for Assessing Response Bias in Questionnaire Measures." *Structural Equation Modeling*, **16**(2), 364–381. `doi:10.1080/10705510902751374`.

Goffin R, Christiansen N (2003). "Correcting Personality Tests for Faking: A Review of Popular Personality Tests and an Initial Survey of Researchers." *International Journal of Selection and Assessment*, **11**(4), 340–344. `doi:10.1111/j.0965-075x.2003.00256.x`.

Hägglund G (1982). "Factor Analysis by Instrumental Variables Methods." *Psychometrika*, **47**(2), 209–222. `doi:10.1007/bf02296276`.

Harris JA (1997). "A Further Evaluation of The Aggression Questionnaire: Issues of Validity and Reliability." *Behaviour Research and Therapy*, **35**(11), 1047–1053. `doi:10.1016/s0005-7967(97)00064-8`.

Holden R (2010). "Social Desirability." In *The Corsini Encyclopedia of Psychology*. John Wiley & Sons.

Kaiser HF (1958). "The Varimax Criterion for Analytic Rotation in Factor Analysis." *Psychometrika*, **23**(3), 187–200. `doi:10.1007/bf02289233`.

Kiers HAL (1994). "Simplimax: Oblique Rotation to an Optimal Target with Simple Structure." *Psychometrika*, **59**(4), 567–579. `doi:10.1007/bf02294392`.

Korth B, Tucker LR (1976). "Procrustes Matching by Congruence Coefficients." *Psychometrika*, **41**(4), 531–535. `doi:10.1007/bf02296973`.

Lorenzo-Seva U (1999). "Promin: A Method for Oblique Factor Rotation." *Multivariate Behavioral Research*, **34**(3), 347–365. `doi:10.1207/s15327906mbr3403_3`.

Lorenzo-Seva U (2003). "A Factor Simplicity Index." *Psychometrika*, **68**(1), 49–60. `doi:10.1007/bf02296652`.

Lorenzo-Seva U, Ferrando PJ (2009). "A General Factor-Analytic Procedure for Assessing Response Bias in Questionnaire Measures." *Structural Equation Modeling*, **16**(2), 364–381. `doi:10.1080/10705510902751374`.

Lorenzo-Seva U, Ferrando PJ (2013). "**FACTOR** 9.2: A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models." *Applied Psychological Measurement*, **37**(6), 497–498. `doi:10.1177/0146621613487794`.

Lorenzo-Seva U, Ginkel JRV (2016). "Multiple Imputation of Missing Values in Exploratory Factor Analysis of Multidimensional Scales: Estimating Latent Trait Scores." *Anales de Psicología*, **32**(2), 596–608. `doi:10.6018/analesps.32.2.215161`.

Lorenzo-Seva U, Rodríguez-Fornells A (2006). "Acquiescent Responding in Balanced Multidimensional Scales and Exploratory Factor Analysis." *Psychometrika*, **71**(4), 769–777. `doi:10.1007/s11336-004-1207-4`.

Mas-Herrero E, Marco-Pallares J, Lorenzo-Seva U, Zatorre RJ, Rodriguez-Fornells A (2013). "Individual Differences in Music Reward Experiences." *Music Perception*, **31**(2), 118–138. `doi:10.1525/mp.2013.31.2.118`.

Mirowsky J, Ross C (1991). "Eliminating Defense and Agreement Bias from Measures of the Sense of Control: A 2×2 Index." *Social Psychology Quarterly*, **54**(2), 127–145. `doi:10.2307/2786931`.

Morales-Vives F, Camps E, Lorenzo-Seva U (2013). "Development and Validation of the Psychological Maturity Assessment Scale (PSYMAS)." *European Journal of Psychological Assessment*, **29**, 12–18.

Navarro-Gonzalez D, Lorenzo-Seva U, Vigil-Colet A (2016). "How Response Bias Affects the Factorial Structure of Personality Self-Reports." *Psicothema*, **28**(4), 465–470. `doi:10.7334/psicothema2016.113`.

Neill JA, Jackson DN (1970). "An Evaluation of Item Selection Strategies in Personality Scale Construction." *Journal of Research in Personality*, **30**(3), 647–661. `doi:10.1177/001316447003000312`.

Paulhus DL (1981). "Control of Social Desirability in Personality Inventories: Principal-Factor Deletion." *Journal of Research in Personality*, **15**(3), 383–388. `doi:10.1016/0092-6566(81)90035-0`.

Paulhus DL (1991). "Measurement and Control of Response Bias." In JP Robinson, PR Shaver, LS Wrightsman (eds.), *Measures of Personality and Social Psychological Attitudes*, pp. 17–59. Academic Press. `doi:10.1016/b978-0-12-590241-0.50006-x`.

Paulhus DL, Vazire S (2005). "The Self-Report Method." In RW Robins, RC Fraley, RF Krueger (eds.), *Handbook of Research Methods in Personality Psychology*. Guilford Press.

Rammstedt B, Farmer R (2013). "The Impact of Acquiescence on the Evaluation of Personality Structure." *Psychological Assessment*, **25**(4), 1137–1145. `doi:10.1037/a0033323`.

Rammstedt B, Kemper B (2011). "Measurement Equivalence of the Big Five: Shedding Further Light on Potential Causes of the Educational Bias." *Journal of Research in Personality*, **45**(1), 121–125. `doi:10.1016/j.jrp.2010.11.006`.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Revelle W (2019). **psych**: *Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston. R package version 1.8.12, URL `https://CRAN.R-project.org/package=psych`.

Ruiz-Pamies M, Lorenzo-Seva U, Morales-Vives F, Cosi S, Vigil-Colet A (2014). "I-DAQ: A New Test to Assess Direct and Indirect Aggression Free of Response Bias." *The Spanish Journal of Psychology*, **17**, 310–315. `doi:10.1017/sjp.2014.43`.

Salgado JF (2005). "Personalidad y Deseabilidad Social en Contextos Organizacionales: Implicaciones para la Práctica de la Psicología del Trabajo y las Organizaciones." *Papeles del Psicólogo*, **26**(92), 310–315.

Soto CJ, John OP, Gosling SD, Potter J (2008). "The Developmental Psychometrics of Big Five Self-Reports: Acquiescence, Factor Structure, Coherence, and Differentiation from Ages 10 to 20." *Journal of Personality and Social Psychology*, **94**(4), 718–737. `doi:10.1037/0022-3514.94.4.718`.

The MathWorks Inc (2017). *MATLAB – The Language of Technical Computing, Version R2017a*. Natick. URL `https://www.mathworks.com/products/matlab/`.

Timmerman ME, Lorenzo-Seva U (2011). "Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis." *Psychological Methods*, **16**(2), 385–408. `doi:10.1037/a0023353`.

Vigil-Colet A, Morales-Vives F, Camps E, Tous J, Lorenzo-Seva U (2013). "Development and Validation of the Overall Personality Assessment Scale (OPERAS)." *Psicothema*, **25**(1), 100–106. `doi:10.7334/psicothema2011.411`.

Vigil-Colet A, Ruiz-Pamies M, Anguiano-Carrasco C, Lorenzo-Seva U (2012). "The Impact of Social Desirability on Psychometric Measures of Aggression." *Psicothema*, **24**(2), 115–128.

Viswesvaran C, Ones DS, Hough LM (2001). "Do Impression Management Scales in Personality Inventories Predict Managerial Job Performance Ratings?" *International Journal of Selection and Assessment*, **9**(4), 277–289. `doi:10.1111/1468-2389.00180`.

Affiliation:

David Navarro-Gonzalez
Department of Psychology
Universitat Rovira i Virgili
Ctra de Valls s/n, 43007 – Tarragona, Spain
E-mail: david.navarro@urv.cat