Revista de Administração Contemporânea
ISSN: 1415-6555
ISSN: 1982-7849
Associação Nacional de Pós-Graduação e Pesquisa em Administração

Rogers, Pablo
Best Practices for Your Exploratory Factor Analysis: A Factor Tutorial
Revista de Administração Contemporânea, vol. 26, no. 6, e210085, 2022
Associação Nacional de Pós-Graduação e Pesquisa em Administração

DOI: https://doi.org/10.1590/1982-7849rac2022210085.en

Available in: https://www.redalyc.org/articulo.oa?id=84072128005

Revista de Administração Contemporânea
Journal of Contemporary Administration

ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO

e-ISSN: 1982-7849

Methodological Article

Best Practices for Your Exploratory Factor Analysis: A Factor Tutorial

Melhores Práticas para sua Análise Fatorial Exploratória: Tutorial no Factor

Pablo Rogers*[1]

ABSTRACT

Context: exploratory factor analysis (EFA) is one of the statistical methods most widely used in administration; however, its current practice coexists with rules of thumb and heuristics given half a century ago. **Objective:** the purpose of this article is to present the best practices and recent recommendations for a typical EFA in administration through a practical solution accessible to researchers. **Methods:** in this sense, in addition to discussing current practices versus recommended practices, a tutorial with real data on Factor is illustrated. The Factor software is still little known in the administration area, but is freeware, easy-to-use (point and click), and powerful. The step-by-step tutorial illustrated in the article, in addition to the discussions raised and an additional example, is also available in the format of tutorial videos. **Conclusion:** through the proposed didactic methodology (article-tutorial + video-tutorial), we encourage researchers/methodologists who have mastered a particular technique to do the same. Specifically about EFA, we hope that the presentation of the Factor software, as a first solution, can transcend the current outdated rules of thumb and heuristics, by making best practices accessible to administration researchers.

Keywords: factor analysis; exploratory factor analysis; common factor analysis; principal component analysis; Factor.

RESUMO

Contexto: a análise fatorial exploratória (AFE) é um dos métodos estatísticos mais utilizados em administração. No entanto, sua prática corrente coexiste com regras de bolso e heurísticas proferidas há meio século. **Objetivo:** o propósito deste artigo é apresentar as melhores práticas e recomendações recentes para uma AFE típica em administração através de uma solução prática acessível aos pesquisadores. **Métodos:** nesse sentido, além de serem discutidas as práticas correntes *versus* as práticas recomendadas, ilustra-se um tutorial com dados reais no Factor, um software ainda pouco conhecido na área de administração, porém *freeware*, fácil de usar (*point and click*) e poderoso. O passo a passo ilustrado no artigo, além das discussões levantadas e de um exemplo adicional, também é disponibilizado no formato de vídeos tutoriais. **Conclusão:** através da metodologia didática proposta (artigo-tutorial + vídeo-tutorial), incentivamos os pesquisadores/metodologistas que dominam alguma técnica particular a fazerem o mesmo. Especificamente sobre a AFE, esperamos que a apresentação do software Factor, como uma primeira solução, possa transcender as regras de bolso e heurísticas correntes ultrapassadas, ao tornar acessíveis as melhores práticas para os pesquisadores da administração.

Palavras-chave: análise fatorial; análise fatorial exploratória; análise de fator comum; análise de componentes principais; Factor.

* Corresponding Author.
1. Universidade Federal de Uberlândia, Faculdade de Gestão e Negócios, Uberlândia, MG, Brazil.

Cite as: Rogers, P. (2022). Best practices for your exploratory factor analysis: A Factor tutorial. *Revista de Administração Contemporânea*, 26(6), e210085. https://doi.org/10.1590/1982-7849rac2022210085.en

Published as Early Access: November 11, 2021.
Assigned to this issue: June 06, 2022.

JEL Code: C000.

Editors-in-chief: Wesley Mendes-da-Silva (Fundação Getulio Vargas, EAESP, Brazil)
Marcelo de Souza Bispo (Universidade Federal da Paraíba, PPGA, Brazil)
Associate Editor: Henrique Castro Martins (Fundação Getulio Vargas, EAESP, Brazil)
Reviewers: Jorge Brantes Ferreira (Pontifícia Universidade Católica do Rio de Janeiro, IAG, Brazil)
One of the reviewers did not authorize the disclosure of his/her identity.
Peer Review Report: The disclosure of the Peer Review Report was not authorized by its reviewers.

Received: March 15, 2021
Last version received: June 14, 2021
Accepted: June 30, 2021

# of invited reviewers until the decision:

	1	2	3	4	5	6	7	8	9
1st round	✗	✓	✓						
2nd round	✓	✗							

CONTEXTUALIZATION

Factor analysis (FA) is an interdependence statistical technique that seeks to determine the number and nature of latent variables or factors that explain the variation and covariation in a set of observed measures (Brown, 2015). It has become one of the most widely used multivariate statistical procedures in applied social sciences, particularly psychology, education, sociology, business administration, political science, and public health (Brown, 2015). In applied research, FA is commonly employed to evaluate instruments with multiple items, construct validation, and data reduction (Brown, 2015; Fabrigar & Wegener, 2012). The use of FA in applied social sciences for the purposes mentioned above stems from the common factor model (Brown, 2015; Fabrigar & Wegener, 2012); therefore, it is also known in the literature as common factor analysis (CoFA).

There are two types of CoFA: confirmatory factor analysis (CFA) and explanatory factor analysis (EFA). Both have the goal of reproducing the relationships observed in a group of items in a smaller number of latent variables. CFA and EFA are fundamentally different due to the number and nature of the a priori specifications and restrictions imposed to the model (Brown, 2015).

EFA is used when the researcher does not have a clear or relatively complete expectation about the structure of relationships. CFA is used when the researcher has a clear prediction about the number of specific factors and measures influenced by the factors (Fabrigar & Wegener, 2012). EFA is a data-based approach, meaning no specification is made a priori regarding the number of factors and relationships between factors/items (Brown, 2015). EFA is commonly adopted as a first step to determine the appropriate number of factors and find out which observed variables are indicators of the latent variables (Brown, 2015).

In CFA, the researcher specifies the number of factors and the structure of relationships between the factors and indicators. A pre-established solution is evaluated in terms of how well it reproduces the correlation matrix of the sample of observed variables (Brown, 2015). Unlike EFA, CFA requires a strong conceptual basis to guide the specification and evaluation of the estimated model; therefore, it is recommended for the later stages of the scale development process and construct validation (Brown, 2015).

Two intermediary and complementary approaches have become more prevalent in the literature in the past decade: partial confirmatory factor analysis (PCFA) and exploratory structural equation modeling (ESEM). The former employs conventional EFA and uses adjustment measures found in CFA (Gignac, 2009; Hoelzle & Meyer, 2013). The latter uses the framework/systematization of structural equation modeling (SEM), in which CFA is incorporated, to compare a set of models and groups (Marsh et al., 2009; 2010; Marsh, Morin, Parker, & Kaur, 2014; Marsh, Guo, Dicke, Parker, & Craven, 2020). According to its methodologists, CFA is not a strictly confirmatory technique, nor is EFA a strictly exploratory technique, which justifies a general and integrated framework incorporating the flexibility of EFA and the elimination of the restrictions of CFA. Discussing CFA, PCFA, and ESEM is outside the scope of this article; we recommend a few references for interested readers (Brown, 2015; Gignac, 2009; Kline, 2016; Marôco, 2014; Marsh et al., 2014; Marsh et al., 2009; Thompson, 2004).

EFA is frequently confused with principal component analysis (PCA). This happens because a common extraction method that is wrongly employed in EFA is the principal components method (Brown, 2015; Gaskin & Happell, 2014; Hauck-Filho & Valentini, 2020). EFA can be used for data reduction and for other purposes, but PCA can only be used for that purpose (Gaskin & Happell, 2014; Hauck-Filho & Valentini, 2020; Lloret, Ferreres, Hernández, & Tomás, 2014). Both seek to reduce a set of indicators/items into fewer variables.

PCA is a more parsimonious technique and can contribute with information for decisions regarding the number of factors to retain. It can be the first step to be taken before EFA, especially when dealing with complex problems. However, most methodologists recommend that EFA be employed when the goal is to identify latent constructs responsible for the variation of observed variables (Fabrigar & Wegener, 2012; Gaskin & Happell, 2014; Watkins, 2018). Some data analysts do not consider PCA to be an FA technique, but rather a technique to summarize many variables into fewer components (Hauck-Filho & Valentini, 2020; Henson & Roberts, 2006).

In PCA, the calculations are relatively more simple (Brown, 2015), and it was disseminated due to its lower need for computation power, becoming the standard extraction method in major commercial software (Costello & Osborne, 2005; Lloret et al., 2014). In fact, Baglin (2014), Costello and Osborne (2005), Izquierdo, Olea and Abad (2014), and Lloret, Ferreres, Hernández and Tomás (2014) have argued that this probably explains the popularity of PCA and the confusion when developing an EFA.

For example, in SPSS, the most used commercial software in the field of applied social sciences, PCA is developed in the menu *Analyze/Dimension reduction/ Factor* through the default choice of the extraction method principal component (PC). Therefore, by leaving the software's default option unchanged, the researcher is carrying out a PCA rather than an EFA, as suggested by

the menu. Because of that, many researchers still wrongly believe that PCA is a type of EFA (Damásio, 2012).

Using the default option in SPSS when the goal is to reduce data is not incorrect; however, EFA is often driven by the software default options for purposes of multi-item instruments and construct validation, leading to incorrect decisions regarding the number of factor definitions (Izquierdo, Olea, & Abad, 2014).

This argument can be proven by searching for 'exploratory factor analysis' on YouTube. Most of the videos shown in the platform are of SPSS applications, and many of them adopt the software's default options, even though the problem is related to obtaining a common factor model in the context of applied social sciences. According to Lloret, Ferreres, Hernández and Tomás (2017), the famous default combination in SPSS (principal components + Kaiser criterion + varimax rotation) is probably the worst way to develop an EFA. Unfortunately, this procedure is still very frequently adopted, as well as many other outdated rules of thumb and heuristics (Gaskin & Happell, 2014).

The secondary objective of this article is to discuss traditional heuristics in the context of EFA and to indicate the best practices based on recent evidence, in line with the major decisions researchers must make when developing an EFA: (a) sample size, (b) correlation matrix, (c) extraction method, (d) number of factors, and (e) factorial rotation.

The main objective of the article is to present a step-by-step process for carrying out an EFA in Factor (Ferrando & Lorenzo-Seva, 2017). Lloret et al. (2017) have shown that Factor is the most complete software for carrying out an EFA; it was especially conceived for that purpose and has the advantage of being freeware. According to the authors, it is the most flexible software and it incorporates the recent best practices for EFA (Lloret, Ferreres, Hernández, & Tomás, 2017). The tutorial developed here uses examples based on real data from a multidimensional scale with characteristics similar to what researchers in applied social sciences frequently encounter.

In addition, this article is intended to be a complete multimedia resource, providing readers with video tutorials about the target subjects: (a) a discussion on the best current practices and recent recommendations for developing an EFA in the context of applied social sciences[1], (b) a tutorial for developing an EFA in Factor based on a multidimensional scale presented in the article[2], and (c) an extra Factor tutorial that is not discussed in the article, using real data from a one-dimensional scale[3].

Although the evolution of personal computers made EFA more accessible for researchers, the quality of EFA practices did not seem to improve (Henson & Roberts, 2006; Izquierdo et al., 2014; Lloret et al., 2017; Lloret et al., 2014; Watkins, 2018). According to Gaskin and Happell (2014) and Watkins (2018), one of the main reasons for that is that researchers tend to emulate existing inadequate publications and to trust the standardization of software that indicate non-ideal methods. A significant part of researches using EFA employs inappropriate methods that were developed in the past century, especially because that is still the default option in major statistics software (Damásio, 2012; Gaskin & Happell, 2014; Lloret et al., 2014). With this multimedia tutorial, we hope to contribute to change this scenario by promoting Factor, a free, easy-to-use, point-and-click (Damásio & Dutra, 2018), and powerful (Baglin, 2014; Izquierdo et al., 2014) software with great potential (Lloret et al., 2017). Factor is underused in the field of business administration in Brazil and does not have an accessible instruction manual (Lloret et al., 2017). The authors of the software only added help menus with a few technical notes in its last release (April 2021).

Reviews of professional literature have consistently found many applications of EFA with imprudent method choices and incomplete reports (Watkins, 2018). In the past three decades, systematic literature reviews on the topic of EFA (Bido, Mantovani, & Cohen, 2018; Fabrigar, Wegener, Maccallum, & Strahan, 1999; Ford, Maccallum, & Talt, 1986; Gaskin & Happell, 2014; Henson & Roberts, 2006; Howard, 2016; Izquierdo et al., 2014; Lloret et al., 2017) have reached similar conclusions: there have been high rates of incorrect or unjustified decisions, usually based on classic or out-of-date criteria or defaults.

In that sense, we also intend to contribute to a critical discussion about these classic/out-of-date defaults and to indicate recommendations of best practices in conducting EFA. Unfortunately, in the recent context of business administration in Brazil, imprudent method choices are still adopted and taught (Dias, Silva-Spineli, & Macedo, 2019; Hair, Gabriel, Silva, & Braga, 2019; Vidal, 2016), while major recent advancements in proposals dedicated to EFA are not (Matos & Rodrigues, 2019). This subject will be expanded further in the next section, which contextualizes the major decisions in EFA. We will discuss the example used as a case on Factor based on the following steps: (a) installing the software, (b) loading the database, (c) configuring the analysis, and (d) interpreting the output. At the end, we conclude the tutorial, considering the discussion presented in this introduction and in the next section.

RULES OF THUMB AND BEST PRACTICES IN A TYPICAL EFA

This article does not intend to go into detail about the calculation procedures in EFA or to provide an exhaustive discussion of each of the major decisions

researchers encounter when adjusting an EFA model. For that, we recommend national references indicating major guidelines (Laros, 2012), an intermediate mathematical approach (Mingoti, 2005) and a more advanced mathematical approach (Ferreira, 2018). Unfortunately, to this date, there is no national manual or manual translated into Portuguese incorporating all the recommendations presented in this section. We recommend the following references and/or more detailed texts exclusively about EFA (Brown, 2015; Fabrigar & Wegener, 2012; Thompson, 2004).

The following guidelines and recommendations lead mostly to the typical EFA performed in studies on applied social sciences, that is, those seeking to adjust a common factor model (CoFA) whose measurement scale is at most ordinal, and/or whose data do not show a multivariate normality distribution and/or are strongly skewed (asymmetry outside -1 and +1). These conditions are prevalent in construct validation studies and when evaluating Likert-type multi-item instruments with less than five response options (Baglin, 2014; Lloret et al., 2014).

Regarding sample size, the rules of thumb of minimum size (N), number of observations (n), and amount of parameters (p) propagated in Brazil, especially in the manual by Hair, Black, Babin and Anderson (2014), lack scientific validation. The aforementioned manual recommends an undercut; however, the classic studies cited in it also suggest 20 times larger samples. Suggestions of N = 50, 100, 200, 300, 400, 500, or 1,000 and n:p = 5:1, 10:1, or 20:1 are common in literature (Costello & Osborne, 2005; Gaskin & Happell, 2014; Howard, 2016; Lloret et al., 2014) and meet every requirement with no need for concerns regarding sample size in an EFA.

There is no consensus regarding this point in the literature; however, recent simulation studies agree with the fact that EFA is a procedure for large samples (Costello & Osborne, 2005; Gaskin & Happell, 2014; Lloret et al., 2014). The required sample size is also determined by the nature of data: situations where the researcher anticipates high commonality (>.80), low cross loadings (<.30), and high loadings in each of the factors (>.50) will require smaller samples (Gaskin & Happell, 2014; Lloret et al., 2014); however, these conditions are rare in practice. The number of items per factor and the number of factors may also influence the sample size requirements (Gaskin & Happell, 2014). The more, the better (Costello & Osborne, 2005; Gaskin & Happell, 2014; Howard, 2016; Lloret et al., 2014), but a minimum of 300 observations is recommended for studies dealing with a polychoric matrix (Lloret et al., 2014).

Regarding the data matrix to use, several simulation studies have demonstrated that the Pearson correlation matrix is biased in estimating the population correlation matrix (it underestimates) when data are ordinal (Baglin, 2014; Ferrando & Lorenzo-Seva, 2014; Holgado-Tello, Chacón–Moscoso, Barbero–García, & Vila–Abad 2010; Lloret et al., 2014). The polychoric/tetrachoric correlation matrix is the best option in these cases. The Pearson matrix should only be used for continuous data.

Regarding factor extraction methods, when dealing with ordinal variables, methods based on least squares [OLS (ordinary least squares) family, such as ULS (unweighted LS) and DWLS (diagonally weighted LS)] are the most recommended. Simulation studies indicate that least squares generate the best results in small samples (Gaskin & Happell, 2014; Izquierdo et al., 2014; Lloret et al., 2014). The PC method overestimates the factor loadings and the variance explained by the factors. In addition, when performing FA through PC, the researched is not truly developing a typical EFA (Hauck-Filho & Valentini, 2020). The PC method has been used for decades for EFA problems due to its computational simplicity; however, that argument is no longer valid nowadays (Gaskin & Happell, 2014; Lloret et al., 2014). The maximum likelihood (ML) method is not recommended when using a polychoric matrix. The DWLS (or WLS mean and variance adjusted — WLSMV) method is suggested for violations of multivariate normality and when the variables are ordinal (Izquierdo et al., 2014).

The traditional criteria for factor retention [Kaiser criterion (i.e., eigenvalue > 1), scree plot, and explained variance] have been consistently criticized for overestimating the number of factors (Gaskin & Happell, 2014; Howard, 2016; Izquierdo et al., 2014; Lloret et al., 2014). Any combination of those should be avoided (Lloret et al., 2014). The literature is harsh and emphatic on the use of the Kaiser criterion (the only objective classic criterion) to determine the number of factors. The Kaiser criterion was proposed half a century ago for reasons of computational efficiency. Recent simulation studies do not even consider the Kaiser criterion; they assume it is an inappropriate method and demonstrate that parallel analysis and the hull method perform better, particularly for ordinal data (Gaskin & Happell, 2014; Howard, 2016; Izquierdo et al., 2014; Lloret et al., 2014).

Regarding the rotation methods, oblique rather than orthogonal rotations should be preferred as a standard, since it is more difficult in applied social sciences to justify uncorrelated than correlated factors (Baglin, 2014; Costello & Osborne, 2005; Howard, 2016; Izquierdo et al., 2014; Lloret et al., 2014). In addition, oblique rotations will reproduce an orthogonal rotation when the factors are, in fact, uncorrelated, but the opposite is not true. Among oblique rotations, oblimin and promax are the most frequently used and recommended (Izquierdo et al., 2014). Promin is a simpler and more flexible option available in Factor.

Lastly, regarding the software used to perform a typical EFA, SPSS stands out as the most used commercial

software for applied social sciences. Its default configuration has contributed to the dissemination of the worst decision a researcher can make in an EFA (Lloret et al., 2014; 2017). It allows linear analyses only because it uses only the Pearson correlation matrix. Factor is a free, easy-to-use, and powerful software that encompasses all the recent best practices (Baglin, 2014; Damásio & Dutra, 2018; Ferrando & Lorenzo-Seva, 2017; Lloret et al., 2017). Even with the existing plug-ins for SPSS, Factor is still a better option. If the researcher is familiar with programming lines, another option is the psych package for R.

THE EXAMPLE

The database used in this article as a case study is available in the study by Rogers, Armada and Rogers (2020). It partially used the research instrument, adopting only the WHOQOL-Bref (World Health Organization Quality of Life) scale for our purposes. The scale includes 26 five-point Likert-type items, of which we use only 24. The scale was conceived taking into account four factors (physical, psychological, social, and environmental). However, recent empirical evidence (Perera, Izadikhah, O'Connor, & McIlveen, 2018) supports a more parsimonious (one or two factors) and/or more complex (bifactor or higher order) structure. Rogers et al. (2020) used 918 observations to evaluate the financial well-being versus quality of life and health of Brazilian investors. However, in this study, we did not apply one of the filters used by the authors, developing the example with 1,047 observations.

FACTOR TUTORIAL

Although these are simple procedures, such as installing the software and loading the database, it is important to highlight that some details outlined below may improve the user's productivity due to a few nuances of the program.

Installing the software

Factor is a free software and can be downloaded from the website of the Psychology Department of the Universitat Rovira i Virgili, Tarragona, Spain (Ferrando & Lorenzo-Seva, 2017). Unfortunately, it is currently only available for the Windows operational system. It is a single small file that does not need to be installed. Simply click on the downloaded .exe file to execute the program. Since the file does not take up a lot of computer space, we recommend having the executable file in the same folder as the database for analysis. The user may have as many executable files for the software as he or she deems necessary, depending on the amount of databases being used simultaneously.

Loading the database

After executing the program, we see the window shown in Figure 1a. Click on *Read data* to begin reading the data. The analysis must be configured in the *Configure analysis* menu, and then the *Compute* button will execute the necessary calculations and show the output to be read and interpreted. Figure 1b shows the window where the database can be loaded. We will not discuss the *Multiple group* analysis menu in this article.

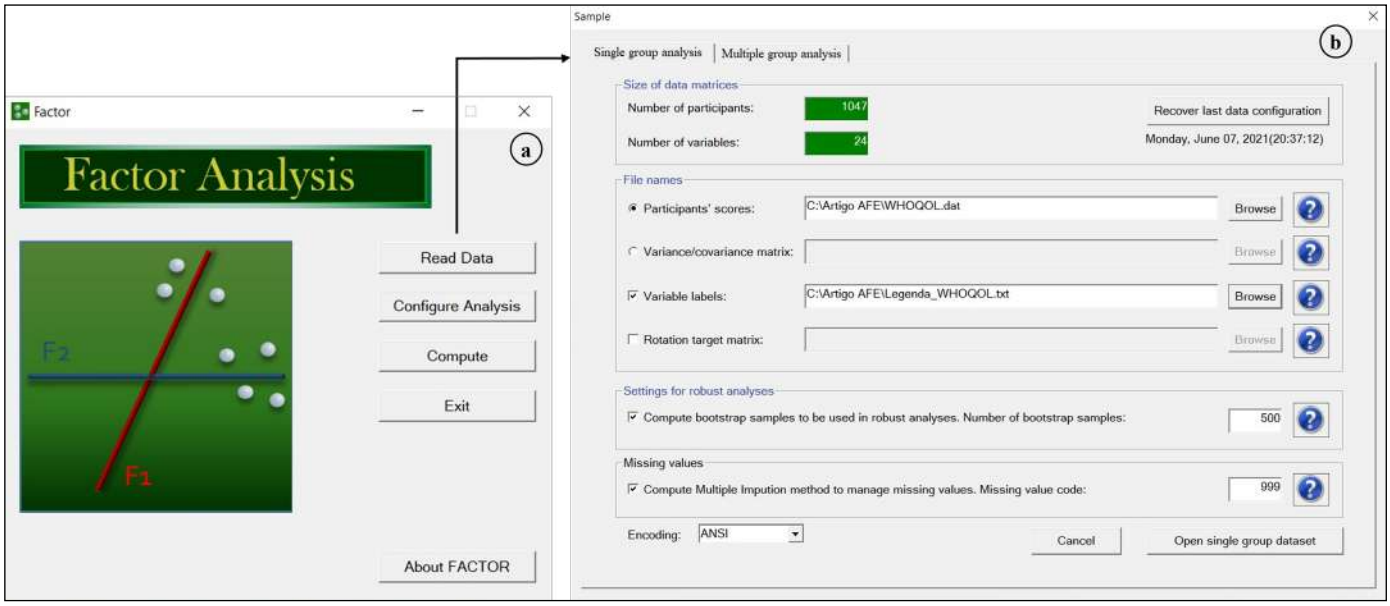


Figure 1. Main window (a) and data read menu (b) in Factor.

The software takes only .txt and .dat files separated by tabs. The user must generate this file structure (save as) in another software, such as Excel or SPSS, since Factor does not have an interface for data editing. Do not include the variables labels in the first line, and include in the file only the variables that will be used for the analysis, as shown in Figure 2a. This is strongly recommended because the software reads variables in a generic format (v1, v2, …, vn), and a large database may decrease productivity and confuse the user. In addition, the software generates bootstrap samples including all the variables in the database before configuring the analyses — any changes in the analysis configuration, such as removing variables, may cause the software to shut down. Optionally, the user can insert a label file for the variables as shown in Figure 2b. This file must have the same number of lines as the number of columns in the database, and a maximum of 40 characters per line.

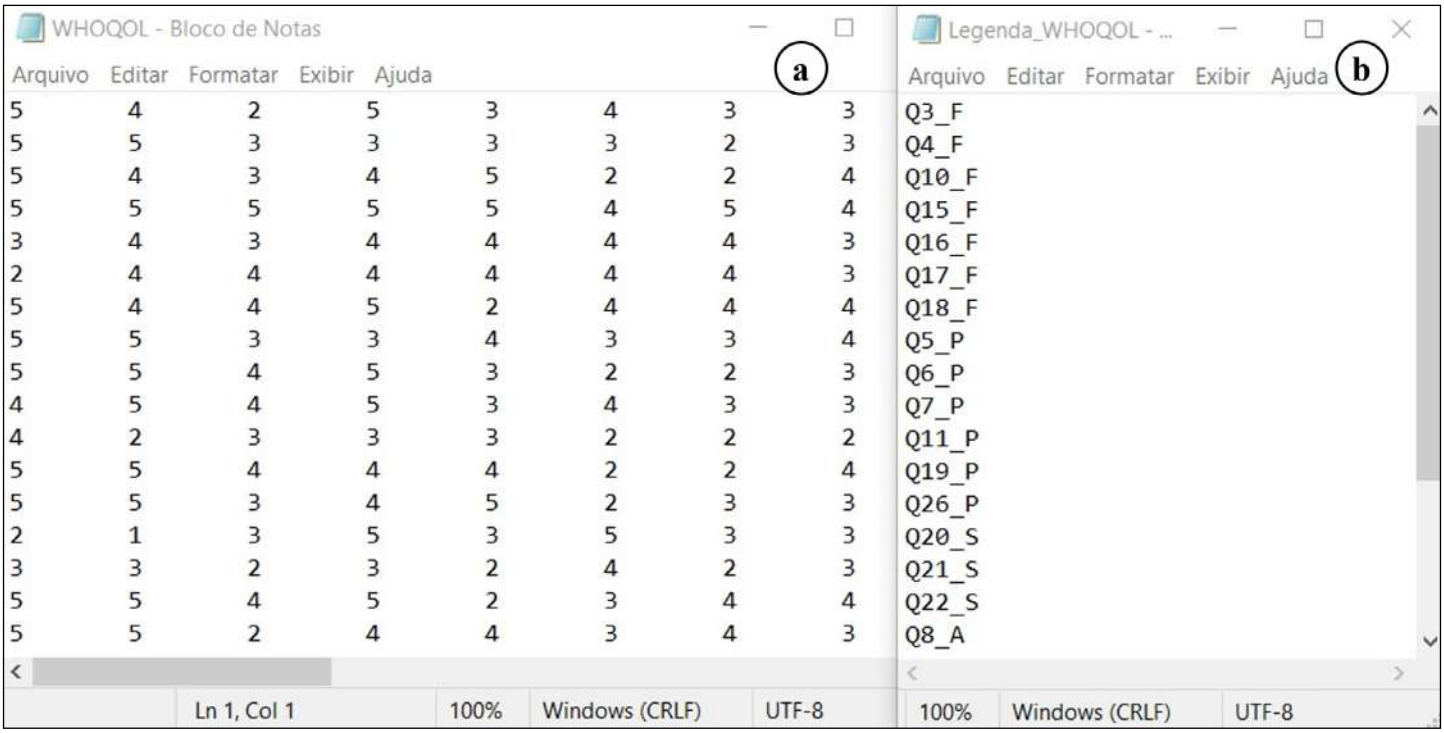


Figure 2. Database (a) and label file (b) in Factor.

If the data was read correctly, the *Numbers of participants* and the *Number of variables* (Figure 1b) will be highlighted in green. Check whether these numbers actually refer to your database. The user could also perform the analyses through a covariance matrix (a square matrix must be loaded in that case). This option is useful when there is no raw data (sourced from the results of an article, for example), and the user wishes to check, replicate, or even compile findings for a meta-analysis. In this data type option, not all adjustment measures are calculated.

If the user anticipates how many factors will be extracted and the weight of the items in each factor, it is possible to define a rotation target matrix a priori for that. If necessary, we encourage readers to click on the question mark button at the side to understand how to configure this matrix. Lastly, we recommend leaving the default selections in Figure 1b: (a) 500 samples are sufficient for robust analysis in *Compute bootstrap* samples, and (b) if there are missing values, insert a code for them in the database. The first selection refers to the procedure to calculate the confidence intervals of the estimates that will be presented in the results through a re-sapling process (bootstrapping).

Do not leave the cells empty and be careful not to indicate the code of a real value. If there is nothing missing in the database, the selection in the menu (*Missing value code*) or the code indicated will not influence the analyses. If there is any empty cell, the entire line will be excluded from the analyses. If there is anything missing identified by the user, the software will manage the missing values based on the proposal of Lorenzo-Seva and van Ginkel (2016). Recent simulation studies have found minimal differences between alternative methods of imputing missing values in the context of EFA (Gaskin & Happell, 2014).

Configuring the analysis

After loading the data, click on the button *Configure analysis* in the main menu. The menu will remain open, but the window shown in Figure 3 will open to enable the selection of all the EFA parameters. Because we did not select *Multiple group analysis* in the data loading window (Figure 1b), we do not need to consider the *Data configuration* menu, even though this menu gives the user the opportunity to view the loaded data. First, at the bottom part of the window, indicate the name of the file which will contain the output (Figure 3a). It will be saved in .txt format in the folder containing the executable file of the software. The top left part of the window (*Variables in the analysis*) will not need any adjustment if your database contains only the variables to be used in the analysis. As we mentioned previously, Factor codes variables in a generic format (v1, v2, …, vn), which is why we suggest using a label file and only inserting the variables of interest in the database. The labels will affect only the output.

The user must make one of the major decisions in EFA at the top right corner of the window (*Matrix analyzed*): the correlation matrix to use, which is often relegated to the default option in popular software. In our example, since we are dealing with an ordinal verification scale with only five answer choices, the polychoric matrix is the most appropriate (Figure 3b).

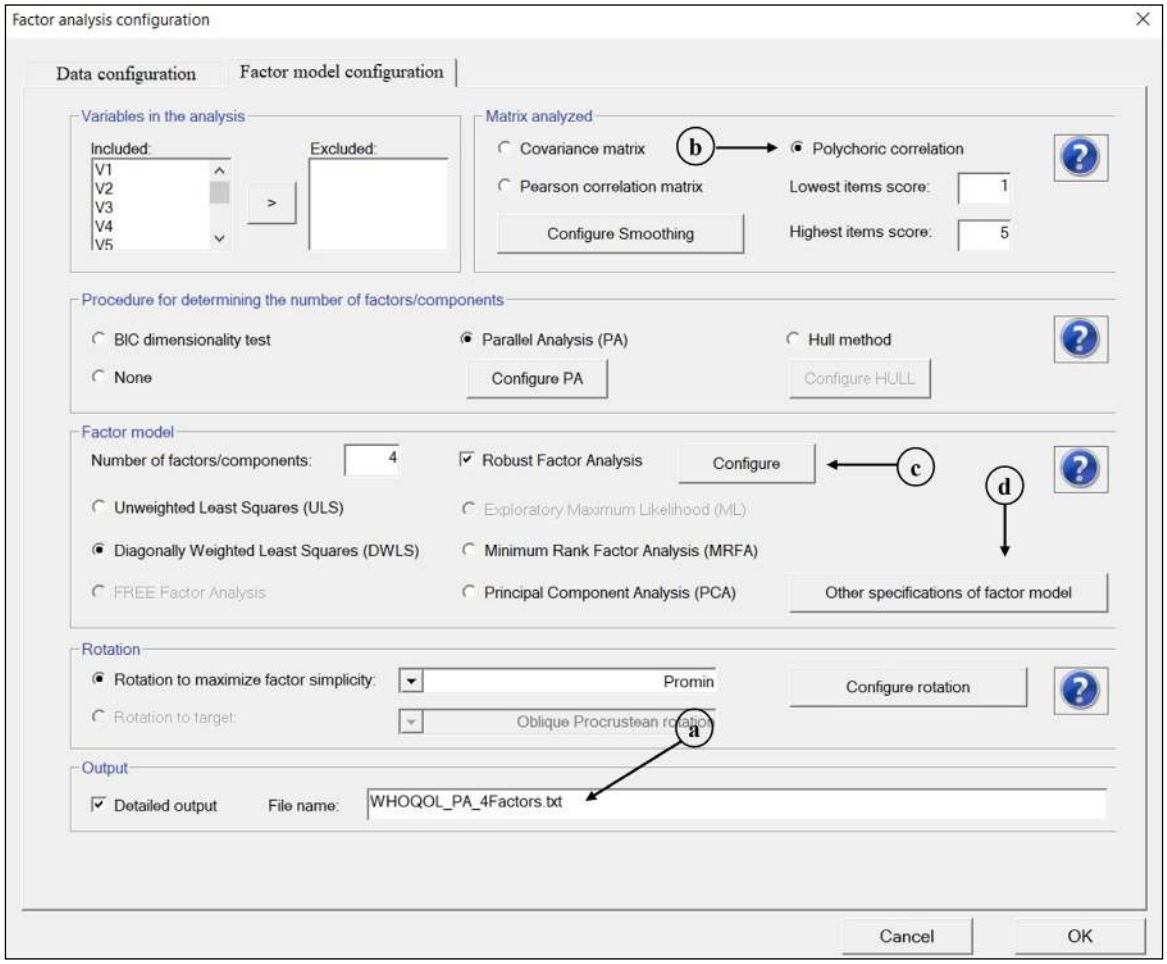


Figure 3. Configuring the analysis in Factor.

In the menus *Configure smoothing* and *Configure PA* (Figure 4), the Factor defaults are usually maintained. The first menu (Figure 4a) refers to the algorithm used to solve the problem of a non-positively defined matrix, which can happen particularly for polychoric matrices when using the least squares methods. In essence, these algorithms consist in smoothing (changing the relative weights) diagonal elements in relation to off-diagonal elements, losing as little variance as possible in the process. The default algorithm in Factor (sweet smoothing) focuses only on the problematic variables, which makes it more efficient at not losing variance (Lorenzo-Seva & Ferrando, 2020).

Factor includes four robust procedures to determine the number of factors: (a) BIC (Bayesian information criterion); (b) minimum average partial (MAP), if the extraction method minimum rank factor analysis (MRFA) is selected; (c) parallel analysis (Figure 4b); and (d) the hull method. Parallel analysis (PA) refers to the solutions of Timmerman and Lorenzo-Seva (2011) and Lorenzo-Seva, Timmerman and Kiers (2011). Horn's classic solution requires a large amount of computational resources. Therefore, Timmerman and Lorenzo-Seva (2011) proposed an optimized implementation of PA that is equally robust, which compares the factors generated to the explained variance instead of the eigenvalues. In this example, we chose optimized PA (Timmerman & Lorenzo-Seva, 2011), as shown in the default options presented in Figure 4b, even though the hull method is equally robust (Lorenzo-Seva & Ferrando, 2019).

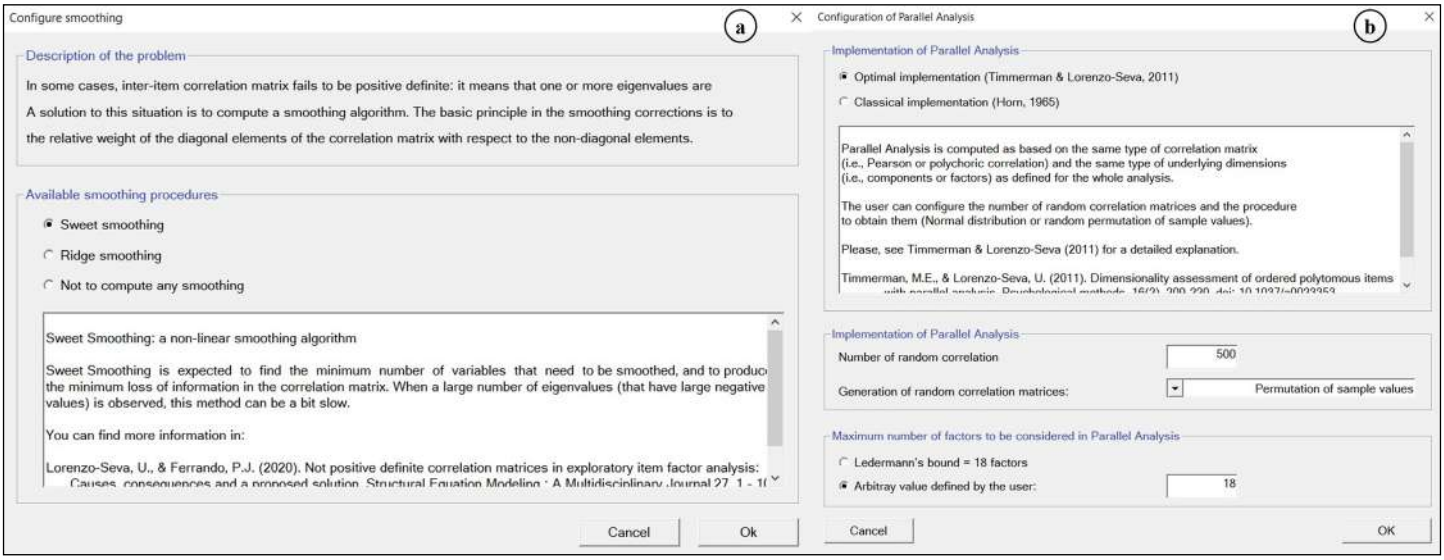


Figure 4. Smoothing (a) and parallel analysis (b) in Factor.

Depending on the number of observations and variables, increasing the number of random matrices from simulations (normal distribution) to the detriment of permutations may greatly increase the calculation time for the analyses. In addition, if *Permutation of sample values* was selected, the correlation matrices will be generated based on the original data, providing more reliable estimates when the observed and populational sample distribution is not normal (Damásio & Dutra, 2018; Timmerman & Lorenzo-Seva, 2011). In fact, it is not recommended to evaluate all the numbers of possible factors (equal to the number of variables) in PA, but at most the number indicated by Ledermann (Timmerman & Lorenzo-Seva, 2011).

In *Factor model* (Figure 3), we defined four factors to estimate due to a priori expectations regarding WHOQOL. The two most recommended methods are ULS and DWLS, since we are dealing with ordinal variables with five ordinal answers. The latter is the same WLSMV used in MPlus. ML will only be available if we do not select a polychoric matrix. The option *FREE factor analysis* is still being developed by the authors of Factor and is not currently available.

Many rotation options are available to the user, as presented in Figure 5. In these conditions, assuming the factors share some degree of relationship, oblique rotations are the most indicated. More specifically, the most used ones are direct oblimin (Figure 5a) and promax (Figure 5b). Unfortunately, the default in most software is an orthogonal rotation (varimax), which should be used only with a good theoretical justification to assume the independence of factors. The default in Factor is promin, an oblique rotation, which results in values very similar to oblimin and is indicated when the extraction method is DWLS (Watkins, 2018). If the user had defined an option for rotation matrix (target weights) in the main menu (*Read data*), the *Rotation target* (Figure 5c) should have been chosen.

The windows that appear when selecting *Configure robust factor analysis* (Figure 3c) and *Other specifications of factor model* (Figure 3d) are expanded upon in Figure 6a and Figures 6b and 6c, respectively.

The first one (Figure 6a) refers to the configurations of the confidence interval estimates via bootstrap; in this case, we are adopting the default option in Factor. The robust method to estimate the adjustment measures is the same used in the MPlus software (WLSMV), which is the most complete when it comes to SEM. As previously mentioned, a sample n = 500 was considered sufficient for our purposes, and the adoption

of the BCa (bias-correlated and accelerated) method is the most recommended to estimate the intervals via bootstrap. The options *Factor/Component loadings values* and *Inter-factor correlations* will compute the bootstrap intervals for the factor loadings and the correlation between factors, respectively. It is important to highlight that Factor is also able to generate models based on the item response theory (IRT) (Damásio & Dutra, 2018).

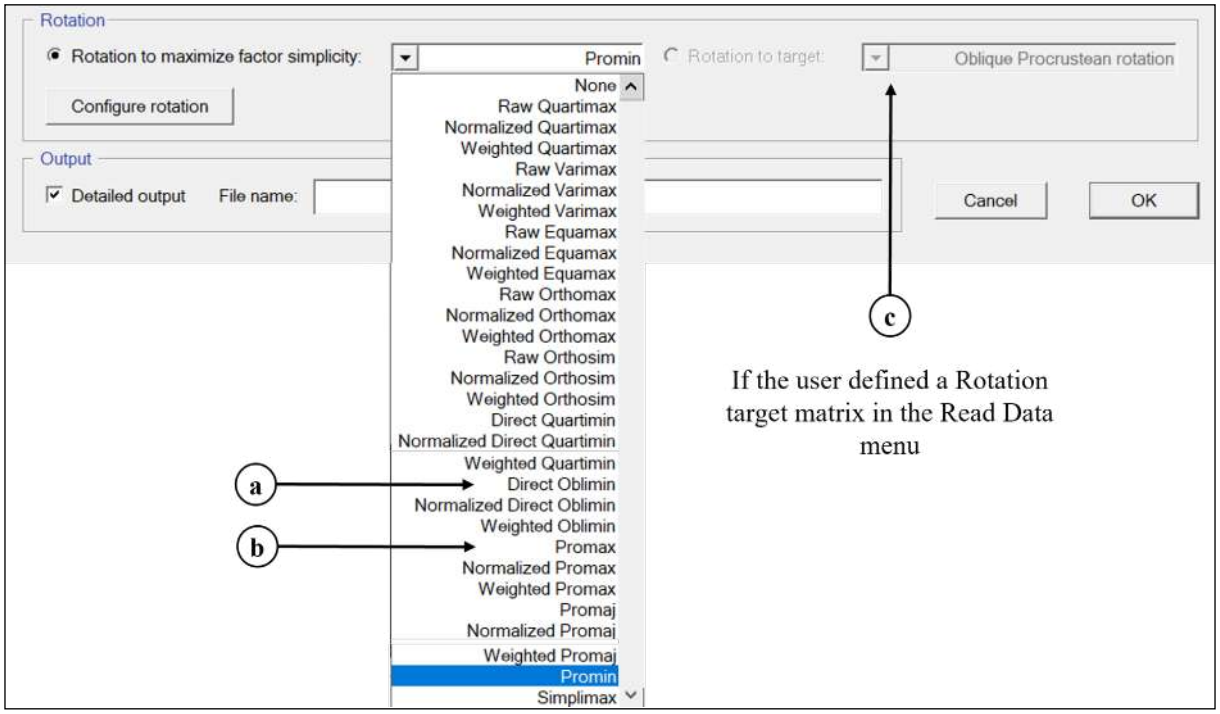


Figure 5. Rotation method in Factor.

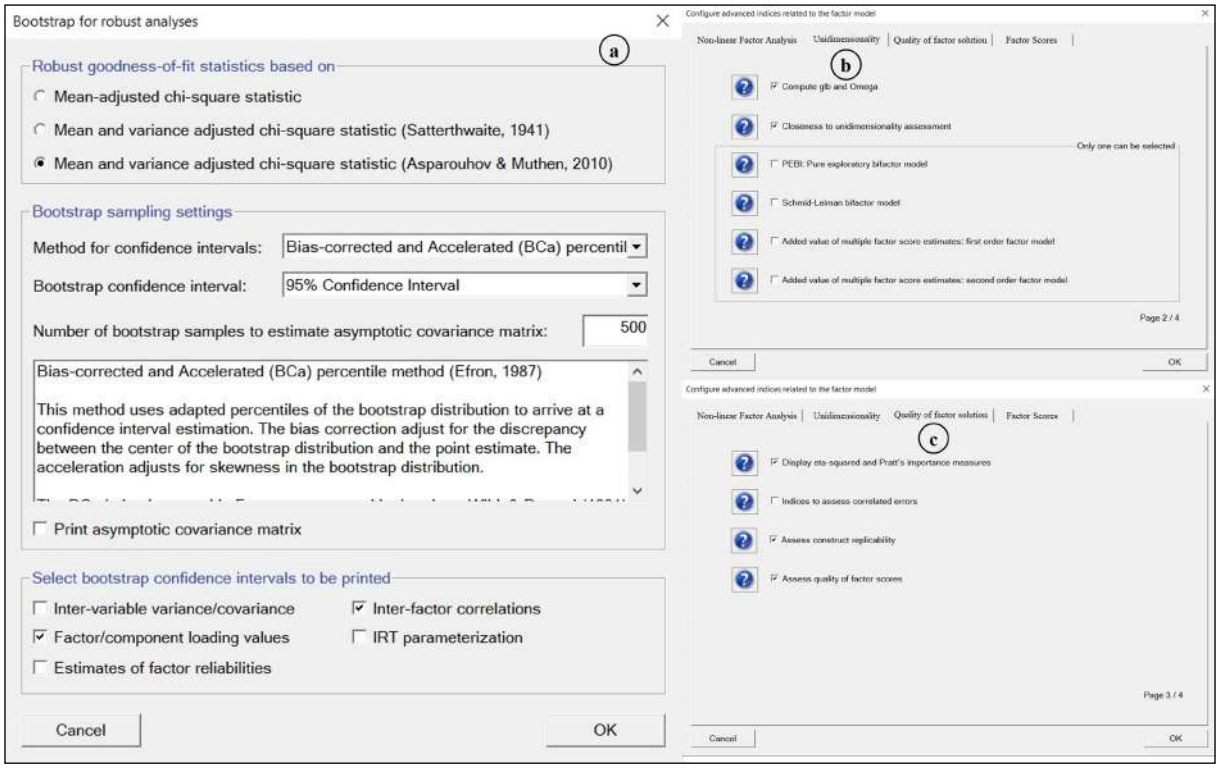


Figure 6. Bootstrap configuration (a) and other specifications of factor model (b and c).

For the other model specifications (Figures 6b and 6c), we will select only the menus *Unidimensionality* and *Quality of factor solutions*. We will not evaluate the parameters (discrimination and difficulty) estimated by an IRT model, and the estimated factor scores will not be necessary. We usually calculate the factor scores when performing the final model adjustment. This way, we can simply indicate the calculation method (a) Bayesian method may be the most appropriate: EAP) and the name of the output file, which will be saved in the folder containing the executable file of the software. In addition to the software default, five other selections are required: (b) *Compute glb and omega* and *Closeness to unidimensionality assessment*, to evaluate the reliability and unidimensionality of the scale; and (c) *Display eta-squared and Pratt's importance measures, Assess construct replicability*, and *Assess quality of factor scores*, to evaluate the effect size of the factors over the items (relevant when examining cross loadings), the replicability of the construct, and the quality of factor scores.

Lastly, *Configure rotation* (Figure 3), if selectable when estimating a multidimensional scale, usually does not need additional input, since it is related to how the rotation algorithm will operate. If any convergence problem arises at first, it may be necessary to increase the number of iterations.

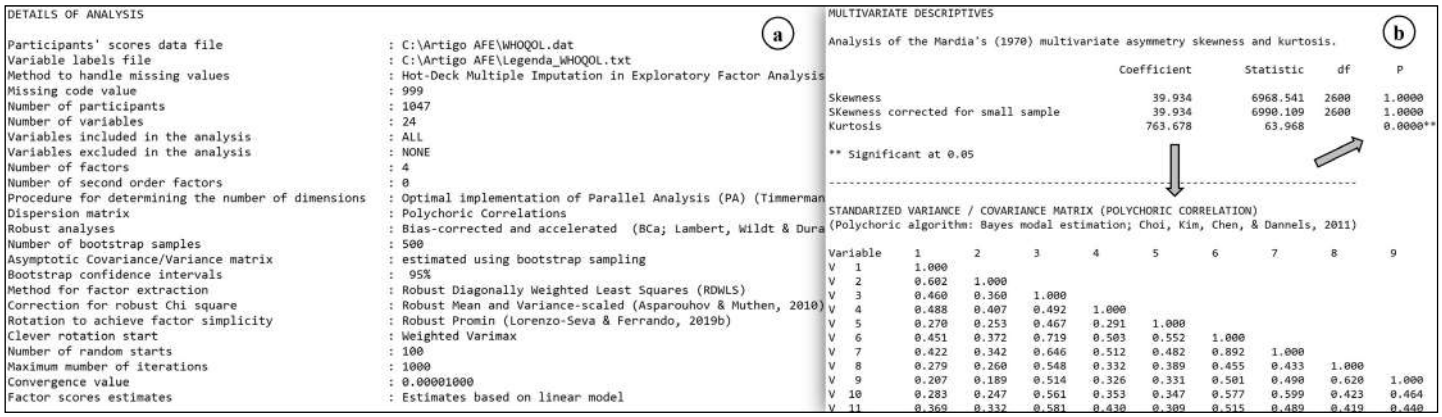
After concluding the EFA configuration, click on *Compute* (Figure 1a). Factor will start calculation the values for the analysis and will open the output when finished. We recommend patience as Factor takes a lot longer than other software in the market regardless of the user's computational power, because it uses a single processor core and its analyses are effectively denser, since the confidence intervals via bootstrap are calculated for almost all measures. This example took 64 minutes to process using a Ryzen 5 and 16 GB of RAM.

Interpreting the outputs

Upon finishing processing, Factor produces a single file with the analyses, which will be opened automatically on notepad or other default software for reading .txt files defined by the user. In this case, it is necessary to scroll down until the end of the file to evaluate each block of interest. We will illustrate and comment on the most important ones following the sequence produced in the output.

The first block in the output (Figure 7a) details the methodological choices configured in the previous step. We recommend looking at these definitions to check for errors made in the configuration step. In addition, if there are problems when computing the polychoric correlation matrix, Factor will automatically use the Pearson matrix. Note that there is a literature citation for each of the choices, and this continues throughout the output, with the references presented at the end. This feature is very important for researchers who wish to reference their methodological choices in scientific studies.

Next, Factor presents some univariate statistics and the bar chart for each of the observed variables entered in the analysis. We will not comment on these blocks. Their interpretation is up to the reader. Next, there are some multivariate statistics (Figure 7b): (a) Mardia's test for multivariate normality, which was not significant in this case, rejecting the possibility that the data present multivariate normality, and (b) the estimated polychoric correlation matrix. Regarding the first finding, the value of multivariate kurtosis is too high to consider the data to be normally distributed, which was expected, given the ordinal nature of the variables. Regarding the correlations, many values under 0.30 for the same variable may indicate lack of adherence of that item to the proposed factor structure.



DETAILS OF ANALYSIS (a)

Participants' scores data file	: C:\Artigo AFE\WHOQOL.dat
Variable labels file	: C:\Artigo AFE\Legenda_WHOQOL.txt
Method to handle missing values	: Hot-Deck Multiple Imputation in Exploratory Factor Analysis
Missing code value	: 999
Number of participants	: 1047
Number of variables	: 24
Variables included in the analysis	: ALL
Variables excluded in the analysis	: NONE
Number of factors	: 4
Number of second order factors	: 0
Procedure for determining the number of dimensions	: Optimal implementation of Parallel Analysis (PA) (Timmerman
Dispersion matrix	: Polychoric Correlations
Robust analyses	: Bias-corrected and accelerated (BCa; Lambert, Wildt & Dura
Number of bootstrap samples	: 500
Asymptotic Covariance/Variance matrix	: estimated using bootstrap sampling
Bootstrap confidence intervals	: 95%
Method for factor extraction	: Robust Diagonally Weighted Least Squares (RDWLS)
Correction for robust Chi square	: Robust Mean and Variance-scaled (Asparouhov & Muthen, 2010)
Rotation to achieve factor simplicity	: Robust Promin (Lorenzo-Seva & Ferrando, 2019b)
Clever rotation start	: Weighted Varimax
Number of random starts	: 100
Maximum mumber of iterations	: 1000
Convergence value	: 0.00001000
Factor scores estimates	: Estimates based on linear model

MULTIVARIATE DESCRIPTIVES (b)

Analysis of the Mardia's (1970) multivariate asymmetry skewness and kurtosis.

	Coefficient	Statistic	df	P
Skewness	39.934	6968.541	2600	1.0000
SKewness corrected for small sample	39.934	6990.109	2600	1.0000
Kurtosis	763.678	63.968		0.0000**

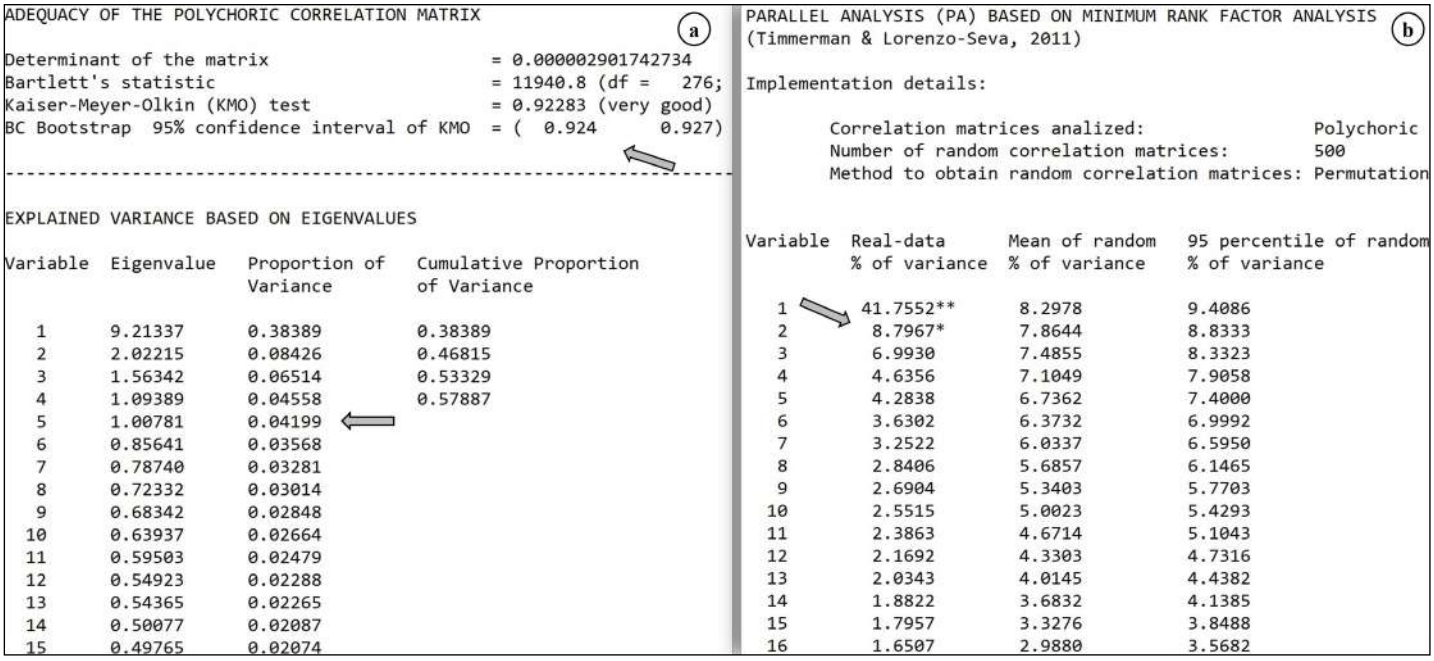
** Significant at 0.05

STANDARIZED VARIANCE / COVARIANCE MATRIX (POLYCHORIC CORRELATION)
(Polychoric algorithm: Bayes modal estimation; Choi, Kim, Chen, & Dannels, 2011)

Variable	1	2	3	4	5	6	7	8	9
V 1	1.000								
V 2	0.602	1.000							
V 3	0.460	0.360	1.000						
V 4	0.488	0.407	0.492	1.000					
V 5	0.270	0.253	0.467	0.291	1.000				
V 6	0.451	0.372	0.719	0.503	0.552	1.000			
V 7	0.422	0.342	0.646	0.512	0.482	0.892	1.000		
V 8	0.279	0.260	0.548	0.332	0.389	0.455	0.433	1.000	
V 9	0.207	0.189	0.514	0.326	0.331	0.501	0.490	0.620	1.000
V 10	0.283	0.247	0.561	0.353	0.347	0.577	0.599	0.423	0.464
V 11	0.369	0.332	0.581	0.430	0.309	0.515	0.489	0.419	0.440

Figure 7. Details of analysis (a) and multivariate statistics (b) in Factor.

Figure 8a shows the first time when the user must interfere in order to evaluate the suitability of the EFA. Bartlett's test was highly significant, rejecting the possibility that the estimated polychoric correlation matrix is equal to an identity matrix. KMO was also very high (0.923), which is above the recommended values for sample suitability (Howard, 2016). These two measures are preliminary and do not say much regarding adjustments of the EFA model. In fact, Barlett's test is not considered important by many authors and has been relegated in recent empirical studies (Ferrando & Lorenzo-Seva, 2018) since it usually rejects the null hypothesis (Howard, 2016).



ADEQUACY OF THE POLYCHORIC CORRELATION MATRIX (a)

```
Determinant of the matrix                    = 0.000002901742734
Bartlett's statistic                         = 11940.8 (df =   276;
Kaiser-Meyer-Olkin (KMO) test                = 0.92283 (very good)
BC Bootstrap  95% confidence interval of KMO = (  0.924      0.927)
```

EXPLAINED VARIANCE BASED ON EIGENVALUES

Variable	Eigenvalue	Proportion of Variance	Cumulative Proportion of Variance
1	9.21337	0.38389	0.38389
2	2.02215	0.08426	0.46815
3	1.56342	0.06514	0.53329
4	1.09389	0.04558	0.57887
5	1.00781	0.04199	
6	0.85641	0.03568	
7	0.78740	0.03281	
8	0.72332	0.03014	
9	0.68342	0.02848	
10	0.63937	0.02664	
11	0.59503	0.02479	
12	0.54923	0.02288	
13	0.54365	0.02265	
14	0.50077	0.02087	
15	0.49765	0.02074	

PARALLEL ANALYSIS (PA) BASED ON MINIMUM RANK FACTOR ANALYSIS (b)
(Timmerman & Lorenzo-Seva, 2011)

Implementation details:

```
Correlation matrices analized:                Polychoric
Number of random correlation matrices:        500
Method to obtain random correlation matrices: Permutation
```

Variable	Real-data % of variance	Mean of random % of variance	95 percentile of random % of variance
1	41.7552**	8.2978	9.4086
2	8.7967*	7.8644	8.8333
3	6.9930	7.4855	8.3323
4	4.6356	7.1049	7.9058
5	4.2838	6.7362	7.4000
6	3.6302	6.3732	6.9992
7	3.2522	6.0337	6.5950
8	2.8406	5.6857	6.1465
9	2.6904	5.3403	5.7703
10	2.5515	5.0023	5.4293
11	2.3863	4.6714	5.1043
12	2.1692	4.3303	4.7316
13	2.0343	4.0145	4.4382
14	1.8822	3.6832	4.1385
15	1.7957	3.3276	3.8488
16	1.6507	2.9880	3.5682

Figure 8. Adequacy of the correlation matrix, eigenvalues (a), and parallel analysis (b) in Factor.

The result of the Kaiser criterion's recommendation for the number of factors can be seen at the bottom part of Figure 8a. In this case, five factors would be chosen (eigenvalue > 1), which at first seems incorrect, since prior knowledge of the factor structure in the WHOQOL scale indicates only four factors. The factor retention technique of the PA (Figure 8b) indicates the selection of only one or two factors, and not four factors as the scale was originally conceived, but it is in line with recent empirical evidence on this measurement model (Perera et al., 2018). According to the PA results, in 500 random matrices generated from the original data, on average, a third factor explained around 7.5% of variance, which is below the 7% explained by the actual data. If a third factor found by chance based on the same data explains more of the variance than a third factor generated by actual data, it is not effectively expected that a third factor exists (i.e., the suggestion of a third factor is likely a sample error). By analogy, we are able to evaluate the other factor retention suggestions.

An important issue from a theoretical point of view (which is often relegated) is the evaluation of unidimensionality. A multifactor structure can be calculated in a single general score, as we often find in empirical studies; however, the opposite cannot happen: an essentially one-dimensional structure cannot be partitioned into subscales. Regarding this issue, readers should go into more detail about the discussion of bifactor or higher order models, which in recent years has helped the discovery of new latent structures (Damásio & Dutra, 2018) and saved many classic measuring instruments (Brown, 2015). Factor offers three measures to evaluate unidimensionality, in line with the proposal by Ferrando and Lorenzo-Seva (2018): (a) one-dimensional congruence (UniCo) > 0.95; (b) explained common variance (ECV) > 0.85; and (c) mean of item residual absolute loadings (Mireal) < 0.30. According to the score values estimated for the example data (Figure 9) using the UniCo and ECA measures, the WHOQOL scale cannot be essentially treated as one-dimensional. However, the Mireal measure indicates that it can be treated as one-dimensional.

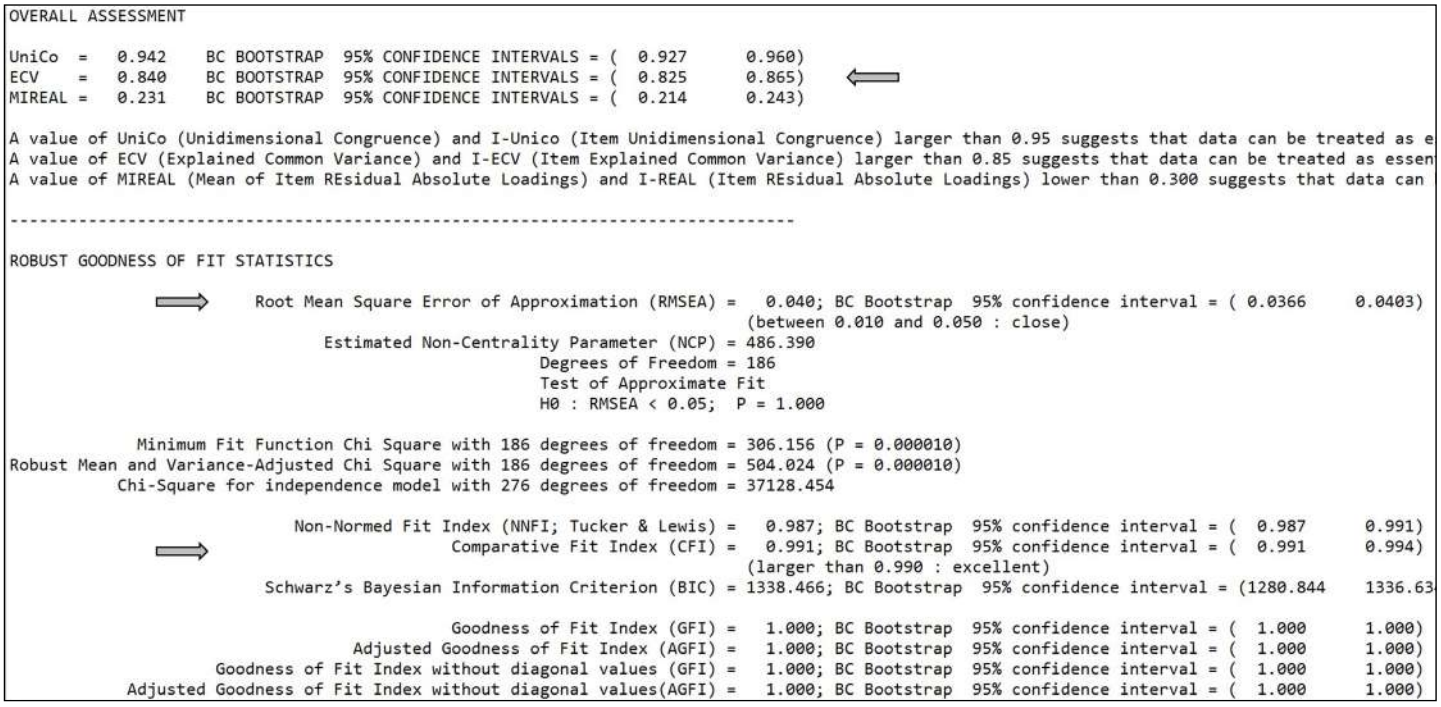


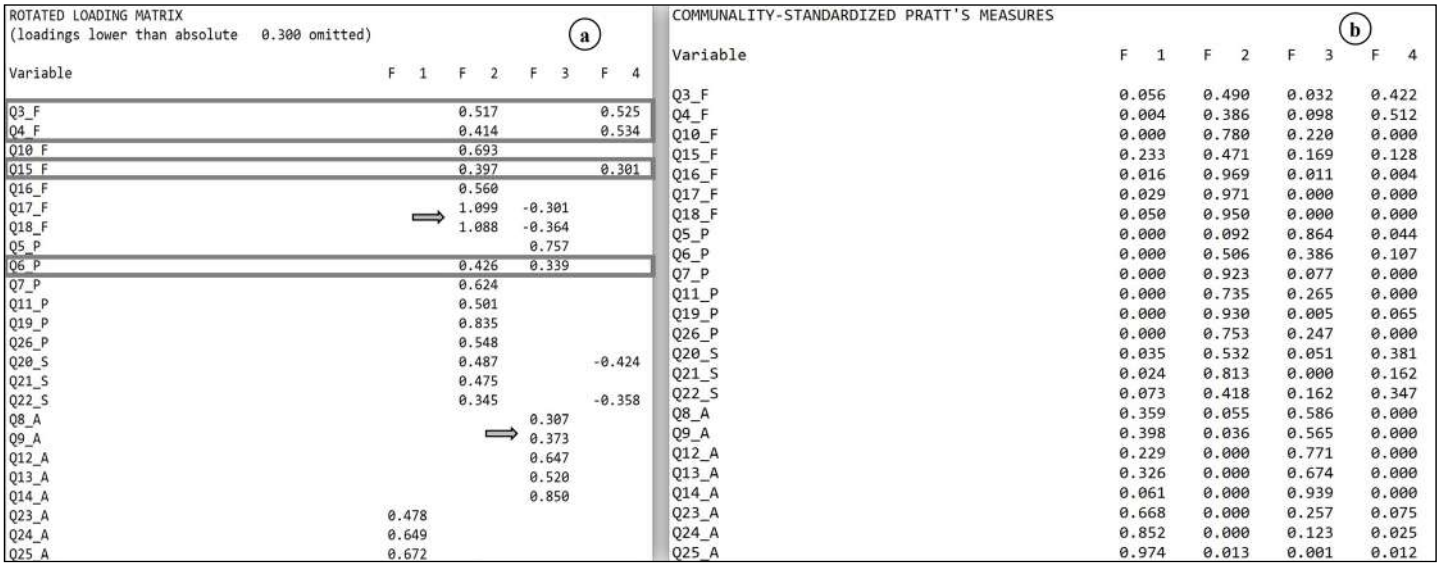
Figure 9. Unidimensionality measures and assessment in Factor.

Figure 9 shows several assessment measures that are usually seen only in CFA. However, they can also be obtained in EFA depending on the factor estimation method. When evaluating the measures indicated in Figure 9 in the context of EFA, we are reproducing a PCFA (Hoelzle & Meyer, 2013). The evaluation of EFA using the criteria required in CFA can be quite rigorous at first (Hoelzle & Meyer, 2013). However, a first look at the main assessment measures of a CFA ($1 < \chi 2/df < 3$, CFI and TLI > 0.95 and RMSEA < 0.08) may indicate how far a simpler or more restricted model will be achieved. Incorporating PCFA when conducting EFA will result in a more sophisticated and insightful understanding of the EFA results (Kline, 2016).

The CFI (0.991) and TLI (NNFI = 0.987) measures seen in Figure 9 indicate an excellent adjustment of the EFA model. The RMSEA measure was also adequate (RMSEA = 0.04; IC95% = 0.037:0.040). The ratio of chi-square to degrees of freedom ($\chi 2/df$ = 486.39/186 = 2.615) also passed the generally accepted threshold. The other measures not highlighted in Figure 9 (GFI and AGFI) have been increasingly relegated by the CFA literature (Damásio & Dutra, 2018; Ferrando & Lorenzo-Seva, 2014; Schreiber, Stage, King, Nora, & Barlow, 2006). However, these measures can be interesting when using a polychoric matrix with some OLS extraction method, since they do not rely directly on the chi-square test (Hancock & Mueller, 2000).

The next block is mostly related to the estimates of the factor loading matrices (unrotated loading matrix, weights of robust rotation, semi-specified target loading matrix, rotated loading matrix, and structure matrix). Since these factor weight estimates are standard outputs in any software running EFA, we chose to show only the rotated factor loadings matrix (>0.30) in Figure 10a in order to carry on with the example. An additional advantage of Factor is that it also shows the factors loadings and the correlation between factors in their confidence intervals via bootstrap (omitted blocks).

Figure 10a evidences several issues with the four factor proposal based on the target data: (a) too many cross loadings, which may suggest a bifactor structure; (b) the items Q17 and Q18, allocated in the physical domain, showed loadings higher than one; and (c) some items with low factor loadings (Q8 and Q9 in the environment domain, for example). There is evidence that the target data does not fit the proposed factor structure. The standardized Pratt's measure (Figure 10b) measures a 'type of effect size' and reaffirms the cross factor loadings found previously, with values close to one indicating that the factor has a high effect on the item, and values close to zero indicating the opposite. We want a single factor to have a high effect on a single item. In fact, to evaluate the simplicity of a factor solution, Factor calculates the loading simplicity (LS) index (omitted output): use this measure to compare factorial weigh matrices between different rotation methods — the closer to one, the better (Lorenzo-Seva, 2003).



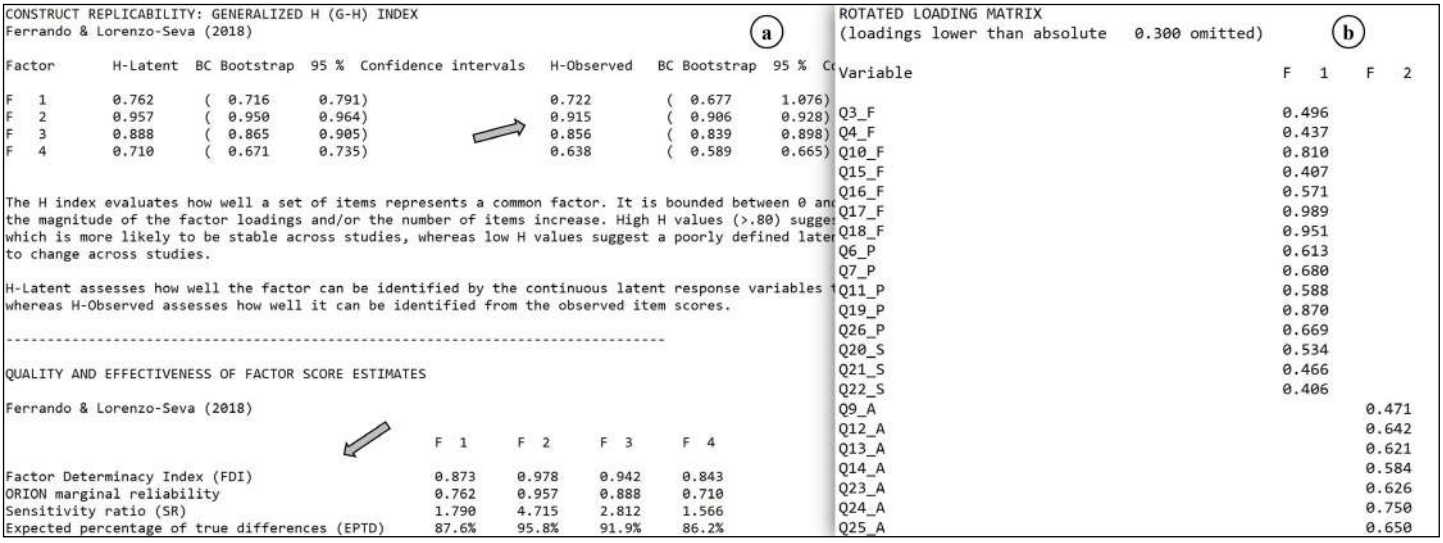
ROTATED LOADING MATRIX (a)
(loadings lower than absolute 0.300 omitted)

Variable	F 1	F 2	F 3	F 4
Q3_F		0.517		0.525
Q4_F		0.414		0.534
Q10_F		0.693		
Q15_F		0.397		0.301
Q16_F		0.560		
Q17_F		1.099	-0.301	
Q18_F		1.088	-0.364	
Q5_P			0.757	
Q6_P		0.426	0.339	
Q7_P		0.624		
Q11_P		0.501		
Q19_P		0.835		
Q26_P		0.548		
Q20_S		0.487		-0.424
Q21_S		0.475		
Q22_S		0.345		-0.358
Q8_A			0.307	
Q9_A			0.373	
Q12_A			0.647	
Q13_A			0.520	
Q14_A			0.850	
Q23_A	0.478			
Q24_A	0.649			
Q25_A	0.672			

COMMUNALITY-STANDARDIZED PRATT'S MEASURES (b)

Variable	F 1	F 2	F 3	F 4
Q3_F	0.056	0.490	0.032	0.422
Q4_F	0.004	0.386	0.098	0.512
Q10_F	0.000	0.780	0.220	0.000
Q15_F	0.233	0.471	0.169	0.128
Q16_F	0.016	0.969	0.011	0.004
Q17_F	0.029	0.971	0.000	0.000
Q18_F	0.050	0.950	0.000	0.000
Q5_P	0.000	0.092	0.864	0.044
Q6_P	0.000	0.506	0.386	0.107
Q7_P	0.000	0.923	0.077	0.000
Q11_P	0.000	0.735	0.265	0.000
Q19_P	0.000	0.930	0.005	0.065
Q26_P	0.000	0.753	0.247	0.000
Q20_S	0.035	0.532	0.051	0.381
Q21_S	0.024	0.813	0.000	0.162
Q22_S	0.073	0.418	0.162	0.347
Q8_A	0.359	0.055	0.586	0.000
Q9_A	0.398	0.036	0.565	0.000
Q12_A	0.229	0.000	0.771	0.000
Q13_A	0.326	0.000	0.674	0.000
Q14_A	0.061	0.000	0.939	0.000
Q23_A	0.668	0.000	0.257	0.075
Q24_A	0.852	0.000	0.123	0.025
Q25_A	0.974	0.013	0.001	0.012

Figure 10. Rotated factor loading matrix (>0.30) (a) and standardized Pratt's measure (b).



CONSTRUCT REPLICABILITY: GENERALIZED H (G-H) INDEX (a)
Ferrando & Lorenzo-Seva (2018)

Factor	H-Latent	BC Bootstrap	95 % Confidence intervals	H-Observed	BC Bootstrap	95 % C
F 1	0.762	(0.716	0.791)	0.722	(0.677	1.076)
F 2	0.957	(0.950	0.964)	0.915	(0.906	0.928)
F 3	0.888	(0.865	0.905)	0.856	(0.839	0.898)
F 4	0.710	(0.671	0.735)	0.638	(0.589	0.665)

The H index evaluates how well a set of items represents a common factor. It is bounded between 0 an
the magnitude of the factor loadings and/or the number of items increase. High H values (>.80) sugge
which is more likely to be stable across studies, whereas low H values suggest a poorly defined late
to change across studies.

H-Latent assesses how well the factor can be identified by the continuous latent response variables
whereas H-Observed assesses how well it can be identified from the observed item scores.

QUALITY AND EFFECTIVENESS OF FACTOR SCORE ESTIMATES

Ferrando & Lorenzo-Seva (2018)

	F 1	F 2	F 3	F 4
Factor Determinacy Index (FDI)	0.873	0.978	0.942	0.843
ORION marginal reliability	0.762	0.957	0.888	0.710
Sensitivity ratio (SR)	1.790	4.715	2.812	1.566
Expected percentage of true differences (EPTD)	87.6%	95.8%	91.9%	86.2%

ROTATED LOADING MATRIX (b)
(loadings lower than absolute 0.300 omitted)

Variable	F 1	F 2
Q3_F	0.496	
Q4_F	0.437	
Q10_F	0.810	
Q15_F	0.407	
Q16_F	0.571	
Q17_F	0.989	
Q18_F	0.951	
Q6_P	0.613	
Q7_P	0.680	
Q11_P	0.588	
Q19_P	0.870	
Q26_P	0.669	
Q20_S	0.534	
Q21_S	0.466	
Q22_S	0.406	
Q9_A		0.471
Q12_A		0.642
Q13_A		0.621
Q14_A		0.584
Q23_A		0.626
Q24_A		0.750
Q25_A		0.650

Figure 11. Replicability and quality measures of the factor scores (a) and final model (b).

In line with the PA results, the block showing a few replicability and quality measures of the factor scores (Figure 11a) indicates a good fit for only two factors (F2 and F3). The first measure, H-observed, is noteworthy. It indicates how much a set of items represents the common factor. Its value ranges from zero to one; values above 0.80 suggest a good definition of the latent variable, which is likely to be more stable for future studies. It is important to highlight that the H-latent index reflects the estimated replicability when the items are interpreted as continuous variables, while the H-observed index reflects the estimated replicability when the items are interpreted as original variables, similar to Likert-type measures (Damásio & Dutra, 2018; Ferrando & Lorenzo-Seva, 2018).

The other measures shown below in Figure 11a (FDI, EAP, SR, and EPTD) are related to the quality of factor scores and whether they can be used for individual evaluations. These measures are recent in the literature and should be evaluated carefully (Ferrando & Lorenzo-Seva, 2018). However, in any case, only the factors F2 and F3 presented a good fit in this research, because: FCI > 0.90; EAP > 0.80, SR > 2; and EPTD > 90% (Ferrando & Lorenzo-Seva, 2018).

Factor also shows the classic reliability measure for the scale (Cronbach's alpha = 0.93), McDonald's omega (ω) (0.93), and the greatest lower bound (GLB = 0.96). Cronbach's alpha has been widely criticized due to the assumption of equivalent weights and a strong influence

of the number of items. McDonald's omega and GLB have been increasingly recommended (Damásio, 2012). At the end of the results file, before the references, Factor also shows information on the residual distribution, which can be interesting to evaluate outliers and missing values, especially if the model is not a good fit.

In an iterative process and after a few rounds, we arrived at the proposal of the factor structure indicated in Figure 11b. It consists of a two-factor model, without the items Q5_P and Q8_A, with the physical, psychological, and social relationships domains in one factor and the environment domain in another factor. This proposal makes more sense from a conceptual point of view, and we have empirical evidence supporting it (Perera et al., 2018). Readers may find themselves in a more demanding position and consider eliminating items with factor loadings under 0.50, for example, leading to a cleaner model to be refined using CFA. We invite readers to watch the video[4] and follow the process until the final model (Figure 11b), including the addition of the comparison between this two-factor model and a bifactor structure. The additional output files until the adjustment of the final model are also available with the article.

CONCLUSION

EFA is one of the most used multivariate interdependence techniques in applied social sciences. Its implementation requires crucial decisions that affect results, and it should not be used seeking the best possible statistical fit to the data; instead, researchers should be grounded on solid theoretical principles to guide its development.

First, if the user is a researcher in business administration, his or her study will potentially involve applying a common factor approach (CoFA). In that sense, it is not about applying PCA due to its simplicity and, in most cases, greater explained variance. PCA seeks to retain as much of the total variance as possible, prioritizing the first factor, and is designed for formative models with the assumption that the items are independent. Studies involving latent factors should seek common variance and reflexive models, predicting the dependency between items. CoFA and PCA are based on different principles, and the choice between one or the other is not statistical, but rather methodological.

For instance, for a user to develop CoFA in SPSS, he or she should not accept the software's default extraction method, as that will result in employing a PCA. Unfortunately, this is one of the first mistakes users make. The two most recommended extraction method options in SPSS for developing a CoFA are ULS and principal axis factoring.

Second, if the study includes variables that are ordinal in nature, the researcher should not use the Pearson correlation matrix. A polychoric matrix is the most indicated, especially if the amplitude of the verification scale is not high, such as between one and five. Pearson's matrix assumes that the relationships are linear and underestimates the strength of the relationships for these types of variables. In this sense, the researcher should immediately discard SPSS, as it still uses only Pearson's matrix. A solution to employ polychoric/tetrachoric matrices would be to use syntaxes developed for SPSS by some authors, such as TETRA-COM (Lorenzo-Seva & Ferrando, 2012) and POLYCOR-C (Lorenzo-Seva & Ferrando, 2015), or R plug-ins in SPSS, such as R-Factor v2.4.3 (Courtney & Gordon, 2013). However, some users may not be familiar with this procedure, and it depends on constant updates of R software packages, including the plug-in itself and SPSS versions, leading to several incompatibility issues.

Third, if the researcher is using a polychoric matrix, DWLS becomes the most suitable method. Factor can perform this type of analysis, including a robust implementation of DWLS, as seen in MPlus. Fourth, the user should not use the Kaiser criterion to decide on the number of factors. This is one of the most crucial decisions in EFA and the literature is emphatic about this methodological choice: eigenvalue > 1, scree plot, and percentage of explained variance are poor choices for factor retention, and the only ones available in SPSS. Simulation studies have consistently proven that PA and the hull method are more accurate. Factor can perform both analyses.

Finally, in a typical EFA, researchers in the field of applied social sciences should use an oblique rotation as the first option for factor interpretation. SPSS serves well in this regard, since the most recommended solutions are available in it: oblimin and promax. However, since the varimax orthogonal rotation is the default in this software and has been recommended for decades for reasons of computational efficiency, it is still the most used in current practice. In research problems in the field of applied social sciences, any orthogonal rotation will hardly have conceptual support.

Unfortunately, researchers tend to use popular rules of thumb rather than evidence-based recommendations when implementing an EFA. Many of these current heuristics were introduced half a century ago, when computational power was limited. This is no longer justifiable today, but these practices continue to spread. This tutorial article sought to contribute to this discussion by criticizing such practices, showing evidence on better options, and presenting a step-by-step solution to help researchers who wish to overcome the barrier of current pragmatism. We hope that this Factor presentation can be a definitive solution in the studies of many researchers in applied social sciences.

NOTES

1. Retrieved from https://youtu.be/ITh1w4tFerA (Accessed on August 13, 2021).
2. Retrieved from https://youtu.be/9X77ARoyys0 (Accessed on August 13, 2021).
3. Retrieved from https://youtu.be/wFTGJG8XRRs (Accessed on August 13, 2021).
4. Retrieved from https://youtu.be/9X77ARoyys0 (Accessed on August 13, 2021).

REFERENCES

Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation*, *19*(5), 1–15. https://psico.fcep.urv.cat/utilitats/factor/

Bido, D. S., Mantovani, D. M. N., & Cohen, E. D. (2018). Destruction of measurement scale through exploratory factor analysis in production and operations research. *Gestão & Produção, 25*(2), 384–397. https://doi.org/10.1590/0104-530X3391-16

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York, NY: The Guilford Press.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Research, and Evaluation Practical Assessment, Research, and Evaluation*, *10*, 7. https://doi.org/10.7275/jyj1-4868

Courtney, M., & Gordon, R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment, Research and Evaluation*, *18*(8), 1–14. https://doi.org/10.7275/9cf5-2m72

Damásio, B. F. (2012). Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica*, *11*(2), 213–228. Retrieved from http://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S1677-04712012000200007

Damásio, B. F., & Dutra, D. de F. (2018). Análise fatorial exploratória: Um tutorial com o Software Factor. In B. F. Damásio & J. C. Borsa (Eds.), *Manual de Desenvolvimento de Instrumentos Psicológicos* (1ª Edição). São Paulo, SP: Vetor Editora.

Dias, P. de S., Silva-Spineli, H. V. R., & Macedo, R. de C. (2019). Estatística multivariadas na Administração: Importância e aplicação da análise fatorial exploratória. *Revista Eletrônica de Administração e Turismo*, *13*(1), 1807–1828. Retrieved from http://www.spell.org.br/documentos/ver/53685/estatisticas-multivariadas-na-administracao--importancia-e-aplicacao-da-analise-fatorial-exploratoria

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford, UK: Oxford University Press.

Fabrigar, L. R., Wegener, D. T., Maccallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in Psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Ferrando, P. J., & Lorenzo-Seva, U. (2014). Exploratory item factor analysis: Additional considerations. *Anales de Psicologia*, *30*(3), 1170–1175. https://doi.org/10.6018/analesps.30.3.199991

Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema*, *29*(2), 236–241. https://doi.org/10.7334/psicothema2016.304

Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, *78*(5), 762–780. https://doi.org/10.1177/0013164417719308

Ferreira, D. F. (2018). *Estatística Multivariada* (3rd ed.). Belo Horizonte, MG: UFMG.

Ford, J. K., Maccallum, R. C., & Talt, M. (1986). The application of exploratory factor analysis in applied Psychology: A critical review and analysis. *Personnel Psychology*, *39*(2), 291–314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x

Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies*, *51*(3), 511–521. https://doi.org/10.1016/j.ijnurstu.2013.10.005

Gignac, G. E. (2009). Partial confirmatory factor analysis: Described and illustrated on the NEO-PI-R. *Journal of Personality Assessment*, *91*(1), 40–47. https://doi.org/10.1080/00223890802484126

Hair, J., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate Data Analysis* (7th ed.). London, UK: Pearson.

Hair, J. F., Jr, Gabriel, M. L. D. S., Silva, D. da, & Braga, S., Junior. (2019). Development and validation of attitudes measurement scales: fundamental and practical aspects. *RAUSP Management Journal, 54*(4), 490–507. https://doi.org/10.1108/RAUSP-05-2019-0098

Hancock, G. R., & Mueller, R. O. (2000). Rethinking construct reliability within latent variable systems. In D. F. S. R. Cudek, & S. H. C. DuToit (Eds.), *Structural equation modeling: Present and future* (pp. 195–216). Lincolnwood, IL: Scientific Software.

Hauck-Filho, N., & Valentini, F. (2020). A análise de componentes principais é útil para selecionar bons itens quando a dimensionalidade dos dados é desconhecida? *Avaliação Psicológica*, *19*(4), A-C. https://doi.org/10.15689/ap.2020.1904.ed

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*(3), 393–416. https://doi.org/10.1177/0013164405282485

Hoelzle, J. B., & Meyer, G. J. (2013). Exploratory factor analysis: Basics and beyond. In *Handbook of Psychology* (2nd ed., pp. 164–188). Hoboke, NJ: John Wiley & Sons.

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, *44*(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, *32*(1), 51–62. https://doi.org/10.1080/10447318.2015.1087664

Izquierdo, I., Olea, J., & Abad, F. J. (2014). El análisis factorial exploratorio en estudios de validación: Usos y recomendaciones. *Psicothema*, *26*(3), 395–400. https://doi.org/10.7334/psicothema2013.349

Kline, R. B. (2016). *Principles and Pratice of Structural Equation Modeling* (4th ed.). New York, NY: The Guilford Press.

Laros, J. A. (2012). O uso da análise fatorial: Algumas diretrizes para pesquisadores. In L. Pasquali (Ed.), *Análise Fatorial para pesquisadores* (pp. 141–160). LabPAM. Retrieved from https://www.researchgate.net/publication/233735561_O_Uso_da_Analise_Fatorial_Algumas_Diretrizes_para_Pesquisadores

Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2014). Exploratory item factor analysis: A practical guide revised and updated. *Anales de Psicologia*, *30*(3), 1151–1169. https://doi.org/10.6018/analesps.30.3.199361

Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2017). The exploratory factor analysis of items: Guided analysis based on empirical data and software. *Anales de Psicología*, *33*(2), 417–432. https://doi.org/10.6018/analesps.33.2.270211

Lorenzo-Seva, U. (2003). A factor simplicity index. *Psychometrika*, *68*(1). https://doi.org/10.1007/BF02296652

Lorenzo-Seva, U., & Ferrando, P. J. (2012). TETRA-COM: A comprehensive SPSS program for estimating the tetrachoric correlation. *Behavior Research Methods*, *44*(4), 1191–1196. https://doi.org/10.3758/s13428-012-0200-6

Lorenzo-Seva, U., & Ferrando, P. J. (2015). POLYMAT-C: a comprehensive SPSS program for computing the polychoric correlation matrix. *Behavior Research Methods*, *47*(3), 884–889. https://doi.org/10.3758/s13428-014-0511-x

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

Lorenzo-Seva, U., & Ferrando, P. J. (2020). Not Positive Definite Correlation Matrices in Exploratory Item Factor Analysis: Causes, Consequences and a Proposed Solution. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(1), 1–10. https://doi.org/10.1080/10705511.2020.1735393

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull Method for Selecting the Number of Common Factors. *Multivariate Behavioral Research*, *46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

Lorenzo-Seva, U., & van Ginkel, J. R. (2016). Multiple Imputation of missing values in exploratory factor analysis of multidimensional scales: estimating latent trait scores. *Anales de Psicología*, *32*(2), 596. https://doi.org/10.6018/analesps.32.2.215161

Marôco, J. (2014). *Análise de equações estruturais: Fundamentos teóricos, software e aplicações* (2nd ed.). Lisbon, PT: ReportNumber.

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

Marsh, H. W., Muthén, B., Morin, Al., Lüdtke, O., Asparouhov, T., & Trautwein, U. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471–491. https://doi.org/10.1037/a0019227

Marsh, Herbert W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set-ESEM: Optimal Balance Between Goodness of Fit and Parsimony. *Multivariate Behavioral Research*, *55*(1), 102–119. https://doi.org/10.1080/00273171.2019.1602503

Marsh, Herbert W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, *16*(3), 439–476. https://doi.org/10.1080/10705510903008220

Matos, S. D. A., & Rodrigues, E. C. (2019). *Análise fatorial*. Brasília, DF: Enap.

Mingoti, S. A. (2005). *Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada*. Belo Horizonte, MG: UFMG.

Perera, H. N., Izadikhah, Z., O'Connor, P., & McIlveen, P. (2018). Resolving Dimensionality Problems With WHOQOL-BREF Item Responses. *Assessment*, *25*(8), 1014–1025. https://doi.org/10.1177/1073191116678925

Rogers, D., Armada, M. J. da R., & Rogers, P. (2020). Bem-estar financeiro, saúde e qualidade de vida de um investidor. *Anais do Encontro da Associação Nacional de Pós-graduação e Pesquisa em Administração*, online, 44.

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. In *Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association. https://doi.org/10.1037/10694-000

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

Vidal, D. F. (2016). Análise fatorial em Administração: Uma aplicação prática com o software SPSS. *Revista Eletrônica Estácio Papirus*, *3*(1), 54–79.

Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Authorship

Pablo Rogers

Universidade Federal de Uberlândia, Faculdade de Gestão e Negócios
Av. João Naves de Ávila, n. 2121, Santa Mônica, 38400-902, Uberlândia, MG, Brazil.

E-mail: pablorogers@ufu.br

https://orcid.org/0000-0002-0093-3834

* Corresponding Author

Funding

The author reported that there is no financial support for the research in this article.

Conflict of Interests

The authors have stated that there is no conflict of interest.

Plagiarism Check

The RAC maintains the practice of submitting all documents approved for publication to the plagiarism check, using specific tools, e.g.: iThenticate.

Peer Review Method

This content was evaluated using the double-blind peer review process. The disclosure of the reviewers' information on the first page, as well as the Peer Review Report, is made only after concluding the evaluation process, and with the voluntary consent of the respective reviewers and authors.

Author's Contributions

1st author: conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); validation (lead); visualization (lead); writing-original draft (lead); writing-review & editing (lead).

Copyrights

Data Availability

RAC advises that the videos and their content mentioned in this article were an option by the author to offer additional material to readers. As such, the videos were not part of the peer review process and are not an official RAC publication.

The authors claim that all data used in the research have been made publicly available through the Harvard Dataverse platform and can be accessed at:

Rogers, Pablo, 2021, "Replication Data for "Best Practices for Your Exploratory Factor Analysis: a Factor Tutorial" published by RAC-Revista de Administração Contemporânea", Harvard Dataverse, V1. https://doi.org/10.7910/DVN/RCX8FF

RAC encourages data sharing but, in compliance with ethical principles, it does not demand the disclosure of any means of identifying research subjects, preserving the privacy of research subjects. The practice of open data is to enable the reproducibility of results, and to ensure the unrestricted transparency of the results of the published research, without requiring the identity of research subjects.