anales de psicología, 2017, vol. 33, nº 2 (may), 417-432
http://dx.doi.org/10.6018/analesps.33.2.270211

ISSN print edition: 0212-9728. ISSN web edition (http://revistas.um.es/analesps): 1695-2294

The exploratory factor analysis of items: guided analysis based on empirical data and software

Susana Lloret, Adoración Ferreres, Ana Hernández, and Inés Tomás

Behavioral Sciences Methodology Department and IDOCAL Research Institute. Universitat de València (Spain)

Título: El análisis factorial exploratorio de los ítems: análisis guiado según los datos empíricos y el software.
Resumen: El objetivo del presente trabajo es ilustrar cómo la aplicación adecuada o inadecuada del análisis factorial exploratorio (AFE) puede llevar a conclusiones muy diferentes. Para ello se evalúa el grado en que cuatro paquetes estadísticos diferentes que permiten realizar AFE de ítems, en concreto SPSS, FACTOR, PRELIS y MPlus, permiten o limitan la aplicación de los estándares actualmente recomendados en materia de análisis factorial. Asimismo se analizan y comparan los resultados que ofrecen dichos programas cuando se factorizan datos empíricos de escalas que ajustan, según el caso, de manera inadecuada, ambigua u óptima a los supuestos del modelo AFE lineal clásico, a través de las distintas posibilidades que ofrecen los distintos programas. Los resultados de la comparación ilustran las consecuencias de elegir entre un programa u otro, y también las consecuencias de elegir entre unas opciones u otras dentro de un mismo programa, en función de la naturaleza de los datos. Finalmente se ofrecen una serie de recomendaciones prácticas dirigidas a los investigadores aplicados con cierta orientación metodológica.
Palabras clave: Análisis Factorial Exploratorio; SPSS; FACTOR; PRELIS; MPlus.

Abstract: The aim of the present study is to illustrate how the appropriate or inappropriate application of exploratory factor analysis (EFA) can lead to quite different conclusions. To reach this goal, we evaluated the degree to which four different programs used to perform an EFA, specifically SPSS, FACTOR, PRELIS and MPlus, allow or limit the application of the currently recommended standards. In addition, we analyze and compare the results offered by the four programs when factor analyzing empirical data from scales that fit the assumptions of the classic linear EFA modeling adequately, ambiguously, or optimally, depending on the case, through the possibilities the different programs offer. The results of the comparison show the consequences of choosing one program or another; and the consequences of selecting some options or others within the same program, depending on the nature of the data. Finally, the study offers practical recommendations for applied researchers with a methodological orientation.
Key words: Exploratory Factor Analysis; SPSS; FACTOR; PRELIS; MPlus.

Introduction

This article is the continuation of "Exploratory factor analysis of items: a revised and updated practical guide" (Lloret, Ferreres, Hernández, & Tomás, 2014), published in this journal. That article and the following one "Exploratory factor analysis of items: some additional considerations" (Ferrando & Lorenzo-Seva, 2014) present the currently recommended standards for the applied researcher in terms of factor analysis (FA). In this second part, first we will review and summarize the degree to which four different statistical packages, SPSS version 22.0, FACTOR version 10.3.01 (Lorenzo-Seva & Ferrando, 2006, 2013), PRELIS[1] version 9.10 (Jöreskog & Sörbom, 2007) and MPlus version 6.12 (Muthén & Muthén, 2007, 1998-2012), allow or limit the application of these standards. Second, we will analyze the results offered by each of these programs when factor analyzing empirical data from scales that inadequately, ambiguously, or optimally fit, depending on the case, the assumptions of the exploratory factor analysis (EFA) classic linear model. These three practical cases with real data will allow us to compare the consequences of choosing a particular software; and the consequences of selecting the different options within the same software, when the data are more or less "problematic". Our objective is clear: to illustrate how the appropriate or inappropriate application of EFA can lead to very different conclusions.

SPSS, FACTOR, PRELIS, and Mplus vary in the degree to which they allow the application of current standards. Following the brief guide we presented in the first part (Lloret et al., 2014), it would be desirable for these programs to offer all the options included in Table 1, or at least most of them. However, not all of them do so.

The information included in Table 1 shows that the most complete program is FACTOR, the only one specifically designed for FA, and a freeware program. It only lacks a factor extraction method such as weighted least squares, and its authors are already working on this (Ferrando & Lorenzo-Seva, 2014). The strength of Mplus lies in the factor extraction methods. However, it leaves out a basic aspect, which is the evaluation of the adequacy of the matrix for factorization, and the researcher is limited by the factor selection criteria Mplus offers. PRELIS limits the researcher even more because it does not allow the user to determine the number of factors to be extracted: it does this automatically by applying the Kaiser criterion (Kaiser, 1958). Finally, a limitation imposed by SPSS is that it only allows the analysis of the items using the linear approach, although this aspect can be mitigated to a certain extent by the use of the free non commercial SPSS programs TETRA-COM and POLYMAT-C (see, Lorenzo-Seva & Ferrando, 2012, 2015). For more information, consult the Annex, which offers a summary of the possibilities and limitations of SPSS, FACTOR, PRELIS and MPlus.

*** Correspondence address [Dirección para correspondencia]:**
Adoración Ferreres Traver. Department of Methodology of Behavioral Sciences. Faculty of Psychology. University of València. Av. Blasco Ibáñez, 21. 46010, Valencia (Spain). E-mail: Adoracion.Ferreres@uv.es

[1] PRELIS is the pre-processor of LISREL.

Table 1. Options allowed by each of the programs analyzed.

	SPSS	FACTOR	PRELIS	MPLUS
TYPE OF DATA				
Analysis of the distribution of	NO	YES	YES	NO
each item	NO	YES	YES	YES
Linear or non-linear approach [a]	YES	YES	NO	NO
Matrix adequacy indicators				
FACTOR EXTRACTION				
Maximum Likelihood (ML)	YES	YES	YES	YES
Robust Maximum Likelihood	NO	NO	NO	YES[b]
Ordinary Least Squares (OLS)	YES	YES	YES[c]	YES
Weighted Least Squares (WLS)	NO	NO	NO	YES
Robust WLS	NO	NO	NO	YES[d]
SELECTION OF NUMBER OF FACTORS				
Based on explained variance:				
Kaiser	YES	YES [e]	YES	YES[e]
Parallel Analysis (PA)	NO	YES	NO	NO
Based on minimization of residuals:				
MAP	NO	YES	NO	NO
RMSR	NO	YES	NO	YES
GFI	NO	YES	NO	NO
Based on goodness of fit:				
Chi-squared	YES	YES	NO*	YES
CFI / NNFI	NO	YES	NO*	YES
RMSEA	NO	YES	NO*	YES
ROTATION METHOD				
Orthogonal	YES	YES	YES	YES
Oblique	YES	YES	YES [f]	YES
Specification of target matrix	NO	YES	NO	YES

Notes: ML = Maximum likelihood; ULS = Unweighted Least Squares; WLS = Weighted Least Squares.
[a] The linear approach analyzes the Pearson correlation matrix; the non-linear approach analyzes the polychoric and tetrachoric correlation matrix, depending on the case.
[b] In Mplus, robust ML methods are MLM and MLMV.
[c] In PRELIS, the analogue method to ULS is MINRES (Jöreskog, 2003).
[d] In Mplus, the robust WLS methods are WLSM and WLSMV.
[e] The program offers information about this criterion although it is not an option to choose; it is offered when ML is used as the estimation method.
[f] PRELIS only offers the pattern matrix in the oblique rotation.

Taking into account the current standards recommended for performing EFA (Ferrando & Lorenzo-Seva, 2014; Lloret et al., 2014) and the options of the different programs (whose main characteristics and strong and weak points are described in the Annex), various recommended roadmaps can be offered for each program. We hope these results are useful for the applied researcher. The recommended road map for the SPSS program is the following: 1) items that show average difficulty and discrimination levels, have a sufficient number of response categories (minimum 5), and present approximately normal distributions; 2) are analyzed by means of Maximum Likelihood (ML), or Unweighted Least Squares (ULS) estimation methods (knowing that ML provides "reliable" indicators of the goodness of fit of the model, although at times non-convergence estimation problems or Heywood[2] cases can appear, whereas ULS is more robust in the estimations (specially on complex solutions) but less robust in the assessment of goodness of fit); 3) combining different factor selection criteria (Kaiser, scree-test, explained variance, and baseline theory); and 4) opting for an oblique rotation method such as PROMAX or OBLIMIN. Nevertheless, researchers must be aware that if the default options of SPSS are used (the famous "Little Jiffy" combination of Principal Components regardless of the type of items + Kaiser criterion + Varimax rotation), it will be probably used in the worst possible way. Unfortunately, this type of EFA analysis by default is still quite frequent.

For FACTOR, we recommend the following two roadmaps depending on the characteristics of the sample and the data. The first would be the one that: 1) fits the non-linear model in large samples, when items are dichotomous or polytomous with few graded response categories, and they do not fulfill the condition of average difficulty and discrimination; 2) analyzing the tetrachoric or polychoric correlation matrix; 3) using the ULS estimation method; 4) combining the different factor selection criteria: Parallel Analysis (PA),the Minimum Average Partial (MAP) test, the goodness of fit indices available when using polychoric correlations, residuals minimization, and, finally, the baseline theory; and 5) opting for the PROMIN rotation, flexible but simple. The other possible combination would be the one that: 1) fits the linear model in small or medium samples, when items show average difficulty and discrimination levels (with approximately normal distributions) and have a sufficient number of response categories (minimum 5); 2) analyzing the Pearson correlation matrix; 3) using ML or ULS estimation methods; 4) combining the different factor selection criteria (PA or MAP, residuals minimization, goodness of fit indices, and the baseline theory); and 5) opting for the PROMIN rotation, flexible but simple. The least recommendable option is to fit the non-linear model using the ULS method when samples are small or questionnaires have a large number of items. In these circumstances, this unweighted method offers problematic solutions in the assessment of fit.

For PRELIS, we recommend following one of these two roadmaps, depending on the characteristics of the sample and the data. 1) For small or medium samples, when the linear model is an adequate approximation to the data because the items have a sufficient number of categories (minimum 5) that reasonably fit a normal distribution; 2) the Pearson correlation matrix is factorized; 3) using the ML or MINRES (MINimum RESiduals) estimation method, equivalent to ULS; 4) following the Kaiser criterion, and evaluating the fit of the successive models that can be compared when employing ML[3]; and 5) the oblique solution is interpreted, which can be compared to the orthogonal solution (both are provided by the program by default). Or 1) for large samples, when the linear model is not appropriate be-

[2] Out-of-range values (e.g., factor loadings greater than one, negative error variances, etc.).

[3] As indicated below, even though, in theory, there is an option to set a certain number of factors to extract, the program always offers the solution suggested by the Kaiser criterion–see detailed information in the Annex.

cause the items are dichotomous or polytomous with few graded response categories that do not fulfill the condition of average difficulty and discrimination; 2) analyzing the tetrachoric or polychoric correlation matrix; 3) using the MINRES estimation method; 4) employing the Kaiser criterion, as it is the only one offered; and 5) interpreting the oblique solution which can be compared to the orthogonal one.

And in MPlus, the two most appropriate road maps, depending on the characteristics of the sample and the data, would be the following: one that 1) analyzes the non-linear model in large samples and dichotomous or polytomous items with few graded response alternatives that do not fulfill the condition of average difficulty and discrimination; 2) analyzing the tetrachoric or polychoric correlation matrix; 3) using a robust Weighted Least Squares (WLS) estimation method such as WLSMV (see details about the characteristics of the different estimation methods in the Annex); 4) combining the different factor selection criteria (residuals minimization, incremental fit indices, and the baseline theory); and 5) opting for an oblique rotation such as GEOMIN. The second route would be to: 1) analyze the linear model in small and medium samples and items with average difficulty and discrimination and a sufficient number of response categories (minimum 5); 2) analyzing the Pearson correlation matrix; 3) using the ML estimation method, or MLMV if the data do not follow an approximately normal distribution; 4) combining the different factor selection criteria (residuals minimization, incremental fit indices, and the baseline theory); and 5) opting for an oblique rotation, such as GEOMIN.

As would be expected, we cannot generate an algorithm that guides the researcher's decisions through the different options mentioned, but we can summarize the rules to follow. The decision begins with the focus of the analysis: linear or non-linear. This decision is controversial because it involves the two conflicting aspects that characterize a good model: simplicity and realism. These two qualities do not go together: the non-linear model is realistic, and the linear model is simple. It is worth mentioning that the simplest approach is often also the most useful one because it offers better fit (Ferrando & Lorenzo-Seva, 2014). However, no simulation studies have been conducted to make recommendations in this regard, especially in the case of the factor estimation methods (Muthén & Asparouhov, 2010). For the moment, we can say that the optimal situation to apply the linear approach is the one that analyzes a set of ordinal items that are close to the continuity assumption because they have a normal multivariate distribution or normal univariate distributions and five or more response categories (Flora, LaBrish, & Chalmers, 2012). In addition, the linear approach is recommended when analyzing a relatively large set of items in a relatively small sample, provided that the items have approximately normal univariate distributions or, at least, symmetric distributions and average difficulty and discrimination levels, and have five or more response alternatives (Ferrando & Anguiano-Carrasco, 2010; Muthén & Kaplan, 1985). When these conditions are not present, the non-linear approach will be more realistic (Ferrando & Lorenzo-Seva, 2013, 2014). We must remember that it is especially important to consider the relationship between the sample size, number of items, and number of response alternatives per item. If there are a large number of items and/or they have five or more response alternatives, no matter how large the sample is, the estimations of the polychoric/tetrachoric correlations will be unstable. We summarize all of these recommendations in Table 2.

Table 2. Factor analysis (FA) approach: Conditions for the linear vs. non-linear model.

	LINEAR FA "SIMPLE" model	NON LINEAR FA "REALISTIC" model
TYPE OF DATA		
Distribution of the items	Univariate normality or average difficulty and symmetric distribution	Free
Inter-item correlations	≤ .50	Free
Number of categories[(a)]	≥ 5	≤ 4
Sample size[(a)]	Small	Medium/large
Number of items[(a)]	Variable	Small
Type of matrix	Pearson	Polychoric/tetrachoric
KMO	> .70	> .70
FACTOR EXTRACTION METHODS	ML, (MLM and MLMV)[(b)] ULS (MINRES)[(c)]	ULS (MINRES)[(c)] (WLS, WLSMV)[(b)]
PROGRAM	SPSS, FACTOR, PRELIS, MPlus	FACTOR, PRELIS, MPlus

Notes: [(a)] These options are connected to each other, so that they must be selected by considering them together.
[(b)] Methods available only in MPLus.
[(c)] MINRES is the analogous method to ULS in PRELIS.

We illustrate the application of the recommended roadmaps for the four aforementioned programs. Specifically we will analyze the factor structure of three sets of scales: the Strength and Flexibility scales of the PSDQ (Physical Self-Description Questionnaire; Marsh, Richards, Johnson, Roche, & Tremayne, 1994); the Self-esteem and Self-concept scales of the PSDQ (Marsh et al., 1994); and the D-48 test (Anstey, 1959. Adapted by the I+D+I Department of TEA Editions, 1996). We analyzed real responses, with the special circumstance that these responses fit the assumptions of the linear or classic EFA model optimally, ambiguously, or inadequately, depending on the scale (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Method

Sample

We used two incidental samples and analyzed three sets of data. The first dataset selected corresponds to the responses of a sample of 914 subjects on the Strength and Flexibility Scales of the PSDQ. Following the criteria presented in Table 2, this set of data would represent the optimal situation, as they acceptably meet the assumptions of the classic or linear factor analysis because: 1) the items present approximately normal distributions -with skewness and kurtosis less than 1 in absolute values-; 2) there are 6 response alternatives; 3) the sample size is large (n = 914), and in addition, the favorable conditions of having 6 items per factor, and only two factors are observed. Moreover, these two factors present a low correlation with each other.

The second set of data chosen corresponds to the responses of a sample of 976 subjects on the Self-esteem and Physical Self-concept scales of the PSDQ. This dataset offers an ambiguous fit because it only partially meets the assumptions of the classic or linear factor analysis, as: 1) it does not fulfill the condition of average location/difficulty (item means range between 4.04 and 5.11 on a response scale from 1 to 6, and the skewness coefficients of 6 of the 14 items are greater than 1 in absolute value); 2) but the inter-item correlations range between .25 and .73, with 74% being less than .50 (M = .42, SD = .13), which indicates that the condition of moderate discrimination is met; 3) the items are answered with 6 response options; and 4) the sample size is large (n = 976). Moreover, favorable circumstances of having six and eight items, respectively, to measure only two factors, are observed.

The third set of data corresponds to the responses of a sample of 499 subjects on the D-48 Test. This test measures the "G" factor of intelligence through the correct/error responses the subjects give on a set of 44 domino-type items with different levels of difficulty. This dataset does not meet the assumptions of the classic or linear factor analysis at all because: 1) it does not fulfill the average location/difficulty condition (the distributions of the items are strongly skewed, some quite easy and others quite difficult); 2) the items are dichotomous; 3) the sample size is large (n = 499); 4), but the test is long, containing 44 items; and 5) the items measure one factor. Furthermore, there is a collinearity problem between two items (probably due to an item duplicated by mistake: their correlation was 1). When we detected this anomaly, we decided to leave it in, in order to detect how each program responded to it.

The recommendation of the classic or linear factor analysis in the first set of data is clear, as it represents the optimal condition for linear analysis. In the second set of data, the recommendation about the type of analysis is not as clear, although the linear approach, simpler and more robust, is the first option for analysis. Here we have the ambiguous condition. In the third set of data, however, the recommendation is the non-linear approach, as it represents the inadequate condition for the linear model.

Procedure

We analyzed each set of data under various conditions:

1) Only with SPSS, the "Little Jiffy" criterion: Principal Components Analysis (PCA) + Kaiser criterion + VARIMAX rotation (VAX). We included this combination because, in spite of not being a factor analysis, and therefore not one of the recommended options, it is still the most popular combination and the one most frequently utilized, perhaps because it appears by default in SPSS (see, Izquierdo, Olea, & Abad, 2014).
2) With the four programs (SPSS, FACTOR, PRELIS and MPlus), we applied the most appropriate approach considering the options offered by each program and the criteria presented (see, Table 1 and Table 2). In cases where there was more than one adequate approach, or more than one adequate option within the same approach, we started by applying the most advisable combination, and we modified it based on the results obtained.

In a reiterative way and in each case, we evaluated each solution according to two criteria or principles: the Statistical Plausibility principle and the Theoretical Credibility principle. On the one hand, we consider a statistical solution to be *plausible* when none of the following appears: Convergence problems, non-positive definite matrices, or Heywood cases. These would be indicators that the solution reached, in spite of being statistically possible, is not plausible, but rather forced. On the other hand, we consider that a solution is *credible* when it offers interpretable results considering the content of the items and the meaning of the factors according to the theory.

Results

Strength and Flexibility Scales. Optimal condition for the linear approach

Table 3 shows the set of exploratory factor analyses carried out by means of the four programs under evaluation following the linear approach (i.e. factor analyzing the Pearson correlation matrix). All of these analyses lead to a plausible and credible result that identifies the two expected factors, except the analysis with MPlus (in version 6, used in the present study).

When we apply "Little Jiffy" with SPSS, we obtain similar results to those obtained also through SPSS with the ML + Kaiser + Oblimin combination, and to those obtained through FACTOR with the ML + 2F + Promin combination: 2 well-defined factors[4]. All of the items loaded above .50 on the expected factor and below .30 on the other one. In both programs, SPSS and FACTOR, the selection criteria of the number of factors used in each case lead to the same number that we had previously established: two factors.

In the case of PRELIS (version 9.10), in spite of having a dialogue box with the option to fix the number of factors, the program does not really allow the researcher to do so. Among the examples tested (those from the present study and from others), if we mark the option of retaining a lower number of factors than the number determined by the Kaiser criterion, the output still shows the solution for the number of factors suggested by Kaiser. If a larger number is specified, the program does not print the results, at least among all the examples tested, and it only shows the distributional analyses of the items. The solution is only printed when the number of factors coincides with what is suggested by the Kaiser criterion. In sum, PRELIS automatically applies the Kaiser rule, and in this case, it retains two factors that are well-defined and consistent with what was expected and with the results of SPSS and FACTOR.

Table 3. Strength and Flexibility Scales. Optimal condition for the linear model.

	KMO	FACTOR RETENTION	NUMBER OF FACTORS, AND COMPOSITION	STATISTICAL PLAUSIBILITY	THEORETICAL CREDIBILITY
SPSS					
(PCA+KAISER+VAX) *Little Jiffy*	.88	KAISER	2 MAJOR (r = .22) 39.7 % and 23.4 %	YES	YES
ADEQUATE CONDITION (ML+KAISER+OBLN)	.88	KAISER, SCREE TEST, THEORY	2 MAJOR (r = .22) 39.7 % and 23.4 %	YES	YES
FACTOR					
(ML+2F+PROMIN)	.88	PA (Optional), NNFI, CFI, GFI, RMSR, THEORY	2 MAJOR (r = .29) 46 % and 26 %	YES	YES
PRELIS					
(ML+KAISER+PROMAX)	------	KAISER[(a)]	2 MAJOR	YES	YES
MPLUS					
(ML+IGFI+GEOMIN)	------	TLI, CFI, RMSEA, SRMR	4 (INADEQUATE): 2 MAJOR 2 MINOR	YES	NO

Notes: PCA = principal components analysis; VAX = varimax rotation; ML = maximum likelihood; OBLN = oblimin rotation; 2F = two factors; IGFI = incremental goodness of fit indices; PA = parallel analysis.
[(a)]PRELIS also offers information about the fit of the different models it estimates with ML estimation methods (from zero factors to the number suggested by the Kaiser criterion).

By contrast, with MPlus, four factors were selected. The comparison of the sequence of models it estimates (a model with zero factors up to the model with nine factors, the maximum allowed) shows that the fit improves as the number of factors increases. However, from five factors on, the solution no longer converges. The 4-factor model presents the best fit (CFI = .998, NNFI = .993, RMSEA = .025, SRMR = .011), according to the comparative incremental fit indices. However, an analysis of both, the structure matrix and the pattern matrix, shows the presence of two major factors and two minor ones. The six items in the strength factor are grouped in the first factor, whereas five of the six flexibility items are grouped in the second factor. The third factor includes only one item of the flexibility scale, and in factor 4, no item presents factor loadings above .40. The factor solution of the 4-factor model is statistically plausible but in reality not very credible. However, the 2-factor model already presents satisfactory goodness of fit indices (CFI = .960, NNFI[5] = .939, RMSEA = .077, SRMR = .025), whereas the 1-factor model does not (CFI = .578, NNFI = .484, RMSEA = .224, SRMR = .185). Analyzing both, the structure matrix and the pattern matrix of the 2-factor model, results show well-defined factors, with the expected item groupings.

Finally, it is worth noting that there were no convergence problems or parameter estimates outside their permissible range. In addition, since the model was not complex, we did not consider it necessary to use ULS and we only tested the ML estimation method.

Self-esteem and Self-concept scales. Ambiguous condition for the linear approach

Table 4 shows the set of EFAs that we performed on this set of data. We analyzed the linear and non-linear approaches when the programs allowed it (all of them except SPSS).

When we apply "Little Jiffy" with SPSS, we obtain results similar to those obtained with the ULS + Kaiser +

[4] However, a prior analysis of this same combination performed with the previous version, FACTOR 9.20, yielded a solution that did not reach convergence. This is an example of how lack of convergence is one of the problems that affect ML, as pointed out above.

[5] Also known as TLI.

Oblimin combination: Two major factors, a minor factor, and a mixed item. Specifically, the six overall physical self-concept items are grouped in the first factor, along with an item from the self-esteem scale (ES6) that is mixed. For the mixed item, the factor loadings on the overall physical self-concept and the self-esteem factors are .55 and .44, respectively, when PC is used, and .50 and .40, respectively, when ULS is used. This tendency of PC to overestimate or at least offer higher factor loadings than the true FA methods has been repeatedly observed throughout the different analyses carried out in the present study and documented in previous studies (e.g. Ferrando & Anguiano-Carrasco, 2010; Izquierdo et al., 2014). On the other hand, six of the eight self-esteem items are grouped in the second factor (including the mixed item). Finally, two other self-esteem items (SE1 and SE5), which are found to be redundant because their wording is quite similar, are grouped in the third factor, with high factor loadings (.81 and .84 with PCA, and .69 and .84 with ULS, for SE1 and SE5, respectively).

Table 4. Self-esteem and Self-concept scales. Ambiguous conditions for the linear model.

	KMO	FACTOR RETENTION	NUMBER OF FACTORS, AND COMPOSITION	STATISTICAL PLAUSIBILITY	THEORETICAL CREDIBILITY
SPSS					
(PCA+KAISER+VAX) *Little Jiffy*	.93	KAISER	2 MAJOR 47 % and 11.2 % 1 MINOR, WITH 2 ITEMS, 8 %	YES NO	YES YES
(ML+KAISER+OBLN) [(a)]	.93	KAISER, SCREE TEST, THEORY	1 HEYWOOD	YES	YES (WITHOUT 2 ITEMS)
(ULS+KAISER+OBLN) [(a)]	.93	KAISER, SCREE TEST, THEORY	2 MAJOR 47 % and 11.2 % 1 MINOR (.47 < r < .61)		
FACTOR					
(ML+2F+PROMIN) [(a)]	.93	MAP NNFI, CFI, GFI, RMSR, THEORY	2 MAJOR (r = .71) 47.5 % and 11.2 % 1 MIXED ITEM	YES	YES
(ULS+2F+PROMIN) [(b)]	.93	PA/MAP GFI, RMSR, THEORY	2 MAJOR (r = .73) 52.7 % and 11.2 % 1 MIXED ITEM	YES	YES
PRELIS					
(ML+KAISER+PROMAX) [(a)]	------	KAISER	1 HEYWOOD	NO	YES (WITHOUT 2 ITEMS)
(MINRES+KAISER+PROMAX) [(b)]	------	KAISER	2 MAJOR 1 MINOR 1 MIXED ITEM	YES	 YES (WITHOUT 2 ITEMS)
MPLUS					
(ML+IGFI+GEOMIN) [(a)]	------	TLI, CFI, RMSEA, SRMR	HEYWOOD	NO	NO
(WLSMV+IGFI+GEOMIN) [(b)]	------		2 MAJOR (r = .91) 1 MINOR 1 MIXED ITEM	YES	YES (WITHOUT 2 ITEMS)

Notes: PCA = principal components analysis; VAX = varimax rotation; ML = maximum likelihood; OBLN = oblimin rotation; ULS = unweighted least squares (equivalent to MINRES); 2F = two factors; IGFI = incremental goodness of fit indices; WLSMV = robust weighted least squares; PA = parallel analysis; MAP = minimum average partial test.
[(a)] Linear approach.
[(b)] Non-linear approach.

Even though the data are not normally distributed, if we use ML with SPSS we would obtain a solution with parameter estimates that are outside their permissible range of values. Specifically, there is a Heywood case (a factor loading of 1.027), so that the solution is not statistically plausible. Although the program does not explicitly refer to this factor loading, it does show the following warning: "one or more communalities greater than 1 have been found during the iterations. The resulting solution should be interpreted with caution". After eliminating mixed item SE6 and one of the redundant items (SE5), the 2-factor model is adequate, although item SE1 (item 1 on the self-esteem scale), which appeared in the minor factor, now presents a marginal factor loading of .35.

FACTOR offers practically the same results using the linear and non-linear approaches: the two expected well-defined factors, although again, SE6 appears as a mixed item. The linear "Pearson + ML + 2F + Promin" combination offers loadings slightly inferior to the analogous "Polychoric + ULS + 2F + Promin" combination. We compared the fit of the 2-factor model from each approach on the criteria available in both cases: GFI and RMSR. GFI is .99 in both approaches and RMSR is slightly better in the linear approach (.046 compared to .048). As Ferrando and Lorenzo-Seva (2014) point out, even in conditions where the non-linear model should fit better, the linear model presents better fit. It must be pointed out that the PA criterion recommends one factor for both, the linear and non-linear approaches; however, when the 1-factor model is tested, GFI and RMSR show inadequate goodness-of-fit, leading to the conclusion that two factors are necessary (which is what the theory indicates). Furthermore, the MAP criterion suggests two factors.

Regarding PRELIS, in the non-linear approach (Polychoric + ULS + Kaiser), three factors are obtained. All of the items on the physical self-concept subscale load on the same factor (with factor loadings between .72 and .86 in the PROMAX solution), but the self-esteem items are split into the other two factors (with factor loadings that range between .35 and .81 in the PROMAX solution). The program also detects that item SE6 is mixed. As occurred with SPSS, by eliminating the mixed item and one of the redundant items (SE5), the 2-factor model becomes adequate, although item SE1, as occurs with SPSS, presents a marginal factor loading equal to .38. When using the available linear approach (Pearson + ML + Kaiser), the program also retains three factors, according to the Kaiser rule. However, a Heywood case is obtained (as the program appropriately indicates), so that we do not continue with the interpretation.

MPlus, with a non-linear approach using the WLSMV robust estimation method, recommended in cases of non-normality like this one (Polychoric + WLSMV), does not present any problem. The goodness-of-fit improves as the number of factors increases. However, from six factors on, the solution no longer converges. Following the criterion of comparing the incremental fit indices, the 3-factor model (CFI = .987, NNFI = .977, RMSEA = .073, SRMR = .021) should be selected. Model 2 shows an acceptable fit (CFI = .943, NNFI = .919, RMSEA = .136, SRMR = .052). However, the improvement in fit of model 3 compared to model 2 is not trivial (ΔCFI = .044, ΔNNFI = .058, ΔRMSEA = -.063); whereas the 4-factor model presents an irrelevant improvement over the 3-factor model (ΔCFI = .006, ΔNNFI = .007, ΔRMSEA = -.012). The one-factor model presents unacceptable fit indices (CFI = .868, NNFI = .844, RMSEA = .190, SRMR = .096). On the other hand, MPlus, with the linear approach "Pearson + ML", apparently has no convergence problems, as there are no advisory or warning messages. The only messages indicate that from a certain number of factors (6 factors), convergence is not reached in the rotation algorithm, and therefore the 6-factor model or any other model with more than six factors is not estimated. However, two Heywood cases appear which means that those results are not statistically plausible. Thus, we do not interpret this solution, only interpreting the results of the non-linear approach.

The analysis of both the structure matrix and the pattern matrix shows that the results from SPSS and PRELIS are repeated: Two major factors and one minor factor are obtained, as well as a mixed item. Again, the results show that when the mixed item (SE6) and one of the redundant items (SE5) are eliminated, the 2-factor model is the most adequate (CFI = .992, NNFI = .987, RMSEA = .061, SRMR = .021), with the factor loading of item SE1 being marginal (.375).

D-48. G Factor Scale. Inadequate condition for the linear approach

We will recall that this set of data is especially problematic from the point of view of factor analysis: the markedly asymmetric distributions of the items, which are generally quite easy or quite difficult, dichotomous, in a sample that is not large enough for the large number of items, and the duplication of one of the items, makes it quite difficult to successfully analyze the set of items. Table 5 shows the results we obtained.

Table 5. D-48. Inadequate condition for the linear model: set of non-linear data.

	KMO	FACTOR RETENTION	NUMBER OF FACTORS, AND COMPOSITION	STATISTICAL PLAUSIBILITY	THEORETICAL CREDIBILITY
SPSS					
(PCA+KAISER+VAX)	NO	KAISER	12	YES	NO
Little Jiffy					
(ULS+KAISER+OBLIMIN) [(a)]	NO	KAISER	12 (2 HEYWOOD)	NO	NO
	NO	FIXED to1	(2 HEYWOOD)	NO	NO
FACTOR					
(ULS+1F+PROMIN) [(b)]	.00 Inadequate	Not possible	Not possible	--	--
PRELIS					
(MINRES+KAISER+PROMAX) [(b)]	------	KAISER	13 (2 HEYWOOD)	NO	NO
MPLUS					
(WLSMV+IGFI+GEOMIN) [(b)]	------	TLI, CFI, RMSEA, SRMR	6	YES	NO

Notes: PCA = principal components analysis; VAX = varimax rotation; ULS = unweighted least squares; 1F = one factor; IGFI = incremental goodness of fit indices; WLSMV = robust weighted least squares.
[(a)] Linear approach.
[(b)] Non-linear approach.

With SPSS, we obtained the "Little Jiffy" solution twice: soliciting KMO or not. We were surprised to find that, if we solicit KMO, the program: 1) does not offer it and does not indicate why; and 2) it does report that the Pearson correlation matrix -the only one it can analyze- is not positive definite. Otherwise, the solution is perfectly plausible. If we do not solicit KMO, the program performs the analysis without indicating any anomalies. The solution it offers is statistically plausible, but lacking in credibility: It identifies 12 components. We should point out that the thirteenth component has an eigenvalue of .983. The Kaiser criterion excludes this factor, but it is clear that this criterion is arbitrary. Of the 12 components, nine have three or more items with factor loadings equal to or greater than .40. The interpretation of the conceptual meaning of these factors is quite confusing. This solution clearly reflects difficulty-factors, where items are grouped together depending on their difficulty (see Ferrando, 1994; Ferrando & Anguiano-Carrasco, 2010).

With the ULS factorization method, even though the matrix is not positive definite, this extraction method can factorize it. However, SPSS indicates that the analyzed matrix has problems. Specifically, it shows the following message: "This matrix is not positive definite. Extraction could not be done". The extraction is skipped, and two Heywood cases appear. All of this is a sign that something is wrong. The solution offered, not very trustworthy as the program warns, once again has 12 factors, of which only five are adequately defined by three or more items. If we set the number of factors at 1, which is the number expected if the test really measures the G factor of general intelligence, the communalities are lower than in the other cases, and the factor loadings identified are also low: of the 44 items, only 10 present factor loadings between .40 and .50, and only 3 between .51 and .55. This unique factor only explains 14.5 % of the variance.

With FACTOR, we applied the non-linear approach. We factorized the tetrachoric correlation matrix along with the ULS estimation method and the PA factor selection method. The program automatically offers the distributions of frequencies of each item, along with the information about the skewness and kurtosis coefficients. We observe that all the items are outside the recommended range. The preliminary information that this analysis provides is sufficient to show that with such asymmetrical distributions in both directions, like those that appear in these data, the matrix cannot be factorized adequately. Then the tetrachoric correlation matrix appears, which is estimated with normality. However, from this point on, it becomes evident to even the less experienced researcher that things are not as they should be. The indicators of the adequacy of the correlation matrix to be factorized (KMO and Bartlett's statistic) offer interpretable values, with the label of "unacceptable" next to them. In addition, instead of the estimations of the eigenvalues and factor loadings, there are symbols that clearly indicate that the analysis was not performed. The program reports -in its own way- that these data cannot be factorized.

When we use PRELIS, the program automatically estimates the bivariate normality tests of the continuous responses that underlie the dichotomous items, and it warns that one of the correlations is equal to 1. It yields a solution with thirteen factors that are not interpretable by content. In all the printed solutions for the different rotations, the program warns of the existence of Heywood cases.

Finally, when we used MPlus, the number of factors was set, as on previous occasions, to the maximum allowed, in order to obtain all the possible factor solutions up to 9 factors. When indicating the categorical nature of the data, the program calculated the tetrachoric correlation matrix, and the analysis was run satisfactorily, showing the warning that there were items with correlations equal to 1. Even so, the program offers the goodness of fit indices for the different models tested. In the factor solutions, there are no Heywood cases, and so it can be stated that this factor extraction method is quite robust for the analysis of categorical items, even with very skewed distributions and a correlation of 1 between two of the items. The one-factor model, theoretically expected, presents a clearly unacceptable fit (CFI = .696, NNFI = .681, RMSEA = .050, SRMR = .216). As the number of factors increases, the fit improves progressively. To decide which model presents the best fit, incremental fit indexes were compared. Based on these criteria, the 6-factor model should be selected (CFI = .972, NNFI = .962, RMSEA = .017, SRMR = .08) because the improvement in fit compared to the 5-factor model is not trivial (ΔCFI = .015, ΔNNFI = .018, ΔRMSEA = .004), whereas the 7-factor model presents an irrelevant improvement over the 6-factor model (ΔCFI = .009, ΔNNFI = .010, ΔRMSEA = .002). The factor solution of the 6-factor model is statistically plausible, although this result does not support the theoretical one-factor model defended by the theory. If we only considered this analysis, the conclusion we would reach is that the D-48 is multidimensional. However, the interpretation of these dimensions would be fairly complicated because the 6 factors identified seem to group the items by their levels of difficulty. This extraction method (WLSMV) makes it possible to more objectively interpret the items that load on each factor because it offers the standard errors for each factor loading estimate and, therefore, gives us an idea of how much they deviate from the hypothetical value of 0. In this analysis, the clue that something is wrong is not found in the statistical results obtained, but rather in their substantive interpretation: from this point of view, the results are simply incongruent with the theory that assumes one single dimension for the G factor.

Discussion

The purpose of this study is not to give recipes for EFA but rather to stimulate critical thinking. Factor analysis is part of our daily lives as researchers in psychology, but we are not yet proficient at using this analytical technique. What can we learn from comparing so many analyses, options, and programs? Above all, to think. Beyond statistical criteria and goodness-of-fit indices, we need to think critically about the more abstract principles of Statistical Plausibility and Theoretical Credibility. We will only present one example of critical thinking for each set of data. We leave the rest to the reader.

The dataset corresponding to the D-48 is extremely difficult to analyze based on both the linear model and the non-linear model. How have the four programs we compared indicated this? It depends on the program. On one extreme FACTOR showed signs that the data were inadequate from the beginning, indicating that the matrix was not suitable for factorization; therefore, it did not factorize it and no numerical results were printed. The program prevents an inappropriate use of the EFA. On the other extreme, Mplus can handle anything, including these "difficult" data. Of course, the robust WLS estimation method (WLSMV) is truly robust. There is no indication that something is wrong with the data, although it does warn about the correlation of 1 between two items. It is the only program that does not present Heywood cases. However, is this good or bad? From our point of view, it is not good because it does not warn researchers about a "difficult" set of data to be successfully factorized. Thus, it can mask a problem that is not statistical, but rather has to do with substantive issues or a bad research design. Just looking at the value reached by KMO or the Bartlett's test of sphericity would be sufficient to stop and think before interpreting the solution, which of course is difficult to interpret.

The conclusion we can draw from the ambiguous dataset is that, as Ferrando and Lorenzo-Seva (2014) recommend, when in doubt, it is necessary to try both approaches: linear and non-linear. However, in addition, solutions with different numbers of factors should be tested: at least the one expected based on the theory, and the one recommended by the (various) criteria considered. When allowed to use the criterion established by default in SPSS, PRELIS and Mplus (Kaiser or comparison of models from 0 to 9 factors) the results suggest that 3 factors are needed to explain the relationships among the items of these two scales (self-esteem and self-concept). FACTOR with PA points to one factor, and with MAP to two factors, and it was the 2-factor model the one that showed the best fit, which agrees with the theory. However, we have to compare the models and their goodness-of-fit. We cannot allow a specific program to decide for us because it will never take into account the theoretical credibility, only the statistical plausibility.

Finally, the first set of data we analyzed, the "friendliest" one, shows that MPlus, by using the model comparison procedure, overestimates the number of factors to be selected, indicating that a 4-factor solution is adequate, which does not make sense. Once again, the solution is statistically plausible, but not at all credible in substantive terms: a factor with only one factor loading and another factor with none are very difficult to interpret. Of course, the experienced researcher will realize that the 2-factor solution is satisfactory, even if it is not the best from the point of view of model fit. However, as we mentioned above, it is necessary to search the program to find out what it offers, and not let it take control. Here we will mention that PRELIS, although offering the same 2-factor solution as SPSS and FACTOR, does not allow the researcher to make any decisions beyond defining the type of approach to use (linear or non-linear). Everything else is decided by the program automatically and cannot be changed. There is no room to explore other options or a different number of factors (especially for some estimation methods).

After all these tests, the authors recommend FACTOR: it is specific and flexible, incorporates the current recommendations for EFA, and it is freely distributed. However, as there is always room for improvement, we would like it to be "friendlier": provide a manual that is more didactic, and would make it possible to read data directly from SPSS or EXCEL. However, the program's web page facilitates an EXCEL application that allows data to be preprocessed in EXCEL and then saved to a file that FACTOR can easily read. The program should also offer messages to the user when performing the analysis, avoiding the impression that the program is blocked, as occurs at times in the XP 10.3.01 version and previous versions. Specifically, version10.3.01 is offered compiled in three modalities: 64-bits, 32-bits and XP. Although in the first two modalities, this problem has already been solved, the users of the XP version, which is older, should be forewarned about this aspect. Furthermore, the 64-bit version manages the memory more efficiently, which makes it possible to conclude analyses that the other modalities would finalize without giving results. Consequently, we recommend using the 64-bit modality. It would also be possible to increase the possibilities of non-linear data analysis, incorporating more robust estimation methods. In this regard, the authors have informed us that a new version of this program (version 10.4.01) will soon be available. This version will include among its improvements the robust ULS estimation method. As we can verify, the authors of FACTOR continue working toward further improving their program, and we would like to thank them for this.

References

Anstey, E. (1959). *Test de Dominós*. Buenos Aires: Paidós.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261-280. doi: 10.1177/014662168801200305

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

Browne, M. W. (1972b). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 207-212. doi: 10.1111/j.2044-8317.1972.tb00492.x
Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36,* 111-150. doi: 10.1207/S15327906MBR3601_05
Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.): *Testing structural equation models* (pp. 136-136). Sage Publications
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 464-504. doi: 10.1080/10705510701301834
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 233-255. doi: 10.1207/S15328007SEM0902_5
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*, 272-299. doi: 10.1037/1082-989X.4.3.272
Ferrando, P. J. (1994). El problema del factor de dificultad: una revisión y algunas consideraciones prácticas [The problem of difficult factor: A revisión and some practice considerations]. *Psicológica, 15*, 275-283.
Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología [The factor analysis as method of research in Psychology]. *Papeles del Psicólogo*, *31*, 18-33.
Ferrando, P. J., & Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory.* Technical Report. Retrieved from http://psico.fcep.urv.es/utilitats/factor/
Ferrando, P. J., & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: algunas consideraciones adicionales [Exploratory item factor analysis: Some additional considerations]. *Anales de Psicología, 30*, 1170-1175.
Flora, D. B., LaBrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Quantitative Psychology and Measurement, 3*, 1-21. doi: 10.3389/fpsyg.2012.00055
Harman, H. H. (1980). *Análisis factorial modern [Modern factor analysis].* Madrid: Saltés.
Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to a simple structure. *British Journal of Mathematical and Statistical Psychology, 17*, 65-70. doi: 10.1111/j.2044-8317.1964.tb00244.x
Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185. doi:10.1007/BF02289447
Izquierdo, I., Olea, J., & Abad, F. J. (2014). El análisis factorial exploratorio en estudios de validación: usos y recomendaciones [Exploratory factors analysis in validation studies: Uses and recommendations]. *Psicothema, 26* (3), 395-400. doi: 10.7334/psicothema2013.349
Jöreskog, K. G. (2002). *Structural equation modeling with ordinal variables using LISREL* (updated in 2004). Technical report. Available on http://www.ams.sunysb.edu/~zhu/tmp/Yue/SEM_brain/covariate/covariate2/SEM%20with%20ordinal%20variables%20using%20LISREL.pdf
Jöreskog, K. G. (2003). *Factor analysis by MINRES*. Technical report. Available on http://www.ssicentral.com/lisrel/techdocs/minres.pdf
Jöreskog, K. G., & Sörbom, D. (2007). *LISREL 8.80.* [Computer Software]. Lincolnwood, IL: Scientific Software International, Inc.
Jöreskog, K. G., Sörbom, D., Du Toit, S., & Du Doit, M. (1999). *LISREL 8: New statistical features.* Chicago: Scientific Software International.
Kaiser, H. F. (1958). The varimax criterion for analytical rotation in factor analysis. *Psychometrika, 23*, 187-200. doi:10.1007/BF02289233
Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36. doi:10.1007/BF02291575
Kiers, H. A. L. (1994). Simplimax: Oblique rotation to an optimal target with simple structure. *Psychometrika, 59*, 567-579. doi:10.1007/BF02294392
Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada [Exploratory item factor analysis: A practical guide revised and updated]. *Anales de Psicología, 30*, 1151-1169. doi: 10.5018/analesps.30.3.199361
Lorenzo-Seva, U. (1999). Promin: a method for oblique factor rotation. *Multivariate Behavioral Research*, *34*, 347-356. doi: 10.1207/S15327906MBR3403_3
Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: a computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, 38*, 88-91. doi:10.3758/BF03192753
Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2. A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied Psychological Measurement, 37,* 497-498. doi: 10.1177/0146621613487794
Lorenzo-Seva, U., & Ferrando, P. J. (2012). TETRA-COM: A comprehensive SPSS program for estimating the tetrachoric correlation. *Behavioral Research, 44*, 1191-1196. doi: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. doi:10.3758/s13428-014-0511-x
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,* 596-608. doi: 10.6018/analesps.32.2.215161
Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519-530. doi: 10.1093/biomet/57.3.519
Marsh, H. W., Richards, G. E., Johnson, S., Roche, S., & Tremayne, P. (1994). Physical Self-Description Questionnaire: Psychometric properties and a multitrait-multimethod analysis of relations to existing instruments. *Journal of Sport and Exercise Psychology, 16*, 270-305.
Muthén, B., & Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189. doi: 10.1111/j.2044.8317.1985.tb00832.x
Muthén, B., & Asparouhov, T. (2010). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods, 17,* 313-335. doi : 10.1037/a0026802
Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.) Los Angeles, CA: Muthén & Muthén.
Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.) Los Angeles, CA: Muthén & Muthén.
O'Connor, B. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396-402. doi:10.3758/BF03200807
Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514. doi:10.1007/BF02296192
Trendafilov, N. (1994). A simple method for procrustean rotation in factor analysis using majorization theory. *Multivariate Behavioral Research, 29*, 385-408. doi: 10.1207/s15327906mbr2904_4
Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321–327. doi:10.1007/BF02293557
Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9, 1-26. doi: 10.1177/014662168500900101

(Article received: 21-02-2016; revised: 05-10-2016; accepted: 11-10-2016)

Annex. Summary of the possibilities and limitations of SPSS, FACTOR, PRELIS and MPlus.

Below we present the possibilities and limitations of each of the most widely used programs in the context of applied research, complementing the most suitable roadmaps according to the new standards. Some programs present more advanced options, but due to space limitations, and given the generalist profile of this study, we will not describe them here.

SPSS

1) Factorial model. Type of data and association matrix

Possibilities: The standard version of SPSS only uses the linear factor analysis approach. The non-linear approach is possible by using additional SPSS programs such as TETRA-COM and POLYMAT-C (Lorenzo-Seva & Ferrando, 2012, 2015). Thus, when we use the standard version, SPSS analyzes the Pearson correlation matrix or the variance-covariance matrix, whether or not this is the suitable option according to the new standards. This program evaluates on demand the suitability of the matrix for its factorization by means of the KMO measure and Bartlett's Sphericity Test.

Limitations: SPSS does not directly analyze polychoric or tetrachoric correlation matrices. Although, as we just pointed out, there are programs that can estimate the polychoric and tetrachoric correlation matrices (Lorenzo-Seva & Ferrando, 2012, 2015), they are not integrated within the factor analysis. SPSS does not offer the researcher a preliminary and automatic analysis of the distribution of the items either.

2) Factor estimation methods

Possibilities: SPSS makes it possible to use some of the recommended factor estimation methods: Unweighted Least Squares, (ULS), Generalized Weighted Least squares, (GWLS), and Maximum Likelihood, (ML), and the most traditional principal axes.

Limitations: The default option is principal components (PC). However, we know that PC is not a factor analysis method, and, therefore, it is currently the least recommended estimation procedure for most psychological applications.

3) Factor selection methods

Possibilities: Different methods based on the amount of explained variance, such as the Kaiser method –default option-, the Scree test, and the proportion of variance explained by each factor. In addition, the researcher can use the option of fixing the number of factors according to his/her hypothesis. Finally, if the ML factor estimation method is used, the chi-square goodness-of-fit index is obtained.

Limitations: It does not offer the most recommendable criteria, such as the goodness-of-fit of the fitted factor model (which would allow us to also compare the fit of different rival models), or "extra" objective criteria such as Parallel Analysis (PA, first described in Horn, 1965) or the MAP (Minimum Average Partial) test (Velicer, 1976). However, it should be pointed out that there are macros for SPSS that can be implemented to carry out these analyses (O'Connor, 2000) (on the author's web page, macros can be downloaded for different programs https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html). SPSS does not offer any method based on residual minimization either.

4) Factor rotation and item assignment methods

Possibilities: SPSS offers an adequate variety of orthogonal and oblique rotation methods: OBLIMIN DIRECT, PROMAX, VARIMAX, EQUAMAX and QUARTIMAX. All of them are guided by the Kaiser principle of factorial simplicity: each item is expected to have a high loading in only one factor. The researcher does not determine what that factor is.

Limitations: In this type of rotation guided by the principle of simplicity, the factorial purity of the measurement instruments is a key issue (Ferrando & Lorenzo-Seva, 2013, 2014).

In summary, SPSS can be used efficiently if we take into account the limitations it presents and take advantage of the possibilities it offers, although the conditions where it is adequate are rather limited (unless the available macros are used to estimate the tetrachoric or polychoric correlation matrices when the non-linear model is the most appropriate).

FACTOR

1) Factorial model. Type of data and association matrix

Possibilities: FACTOR makes it possible to choose between linear and non-linear EFA because it can analyze the Pearson correlation matrix, the variances-covariance matrix and, depending on whether the data are polytomous or dichotomous, the polychoric correlation matrix or the tetrachoric correlation matrix. This program automatically offers various tests of the matrix's adequacy for factorization, among which the KMO measure and Bartlett's sphericity test stand out. In addition, it offers the researcher a preliminary and automatic analysis of the distributions of the items to be analyzed and Mardia's multivariate normality test (1970), so that the selection of linear or non-linear factor analysis is an informed choice.

Moreover, FACTOR 10.3.01 handles missing values through a multiple imputation procedure (Lorenzo-Seva & Van Ginkel, 2016). This is a novel procedure in the context of EFA and it is not available in any other commercial software.

Limitations: FACTOR does not directly read the data from SPSS or Excel. Instead, the data file has to be in ASCII (.dat) format, without labels of variables, number of cases, or any information apart from the item scores. The program website does offer an excel file to preprocess the data and save them in ASCII format. This file was recently updated on the FACTOR web page to better adapt it to the updated format of Excel 2010/13.

2) Factor estimation methods

Possibilities: FACTOR makes it possible to use the common estimation methods ULS and ML. In addition, it incorporates another less well-known method: the minimum range factor analysis. This method allows the researcher to interpret the proportion of common variance explained by each retained factor (see, Lorenzo-Seva & Ferrando, 2006, 2013).

In its latest version (10.3.01), the program is offered in three modalities: 64-bits, 32-bits, and XP. The most efficient version is that of 64-bits because results can be obtained much more rapidly, and larger sets of data can be managed. In fact, this version can finalize analyses that the other versions (32-bits and XP) cannot because they produce a memory error.

Limitations: It does not have robust or weighted least squares (WLS) estimation methods, which are recommendable when the non-linear model is appropriate because they lead to more correct assessments of model fit (as long as the samples are large enough). The next version, 10.4.01, will incorporate these weighted estimation methods.

In another vein, analysis with the 32-bits and XP versions of FACTOR can take longer than with other programs (mainly when performing PA, which involves obtaining and analyzing a large number of random samples through bootstrapping).In addition, in the case of using XP, while performing the analysis, Windows can present an error message, and it may seem that the analysis has stopped and the program is not responding, but it is only a question of waiting. At other times, problems appear, and the program does freeze up. In these cases, it would be helpful to have a warning message from the program indicating that it is necessary to close the program and exit because the analysis cannot continue.

3) Factor selection method

Possibilities it offers: FACTOR requires the researcher to indicate the number of factors to retain. However, it aids in making this decision because it offers objective criteria for this purpose: PA classic implementation, PA optimal implementation (which analyzes the same type of correlation matrix as the one that will be analyzed -Pearson or polychoric-), the MAP test, and the HULL method (which makes it possible to choose between different numbers of factors by combining the principle of parsimony and the goodness-of-fit of the resulting model). Finally, it also offers information about the eigenvalues, so that the user can represent them graphically and apply the Cattell scree-test if desired. Furthermore, when the ML or ULS estimation methods are used (in the latter case, only with polychoric correlations), the program offers different goodness of fit indexes, including the chi-square test, GFI, AGFI, NNFI, CFI, RMSEA, and the non-centrality parameter estimate. Finally, it offers descriptive information about the distribution of residuals.

4) Factor rotation and item assignment methods

Possibilities: Like SPSS, FACTOR offers a variety of traditional orthogonal and oblique rotation methods, including OBLIMIN DIRECT, PROMAX and VARIMAX. FACTOR also includes the OBLIMIN and WEIGHTED VARIMAX procedure and other less well-known methods such as PROMAJ (Trendafilov, 1994). In addition, it offers other new and original methods such as SIMPLIMAX (Kiers, 1994) and PROMIN (Lorenzo-Seva, 1999). SIMPLIMAX is efficient but difficult to use because it requires certain specifications from the researcher (see Ferrando & Lorenzo-Seva, 2014). PROMIN is a special SIMPLIMAX case that does not require previous specifications. As Ferrando and Lorenzo-Seva (2014) explain, PROMIN makes the difficult decisions in the model proposal, so that it is quite close to the exploratory pole, where it is only necessary to specify the number of factors. It is as simple as the classic methods.

FACTOR also allows a more confirmatory approach within the EFA. Among the possible options, the "partially-specified procrustean rotation" makes it possible to propose a target matrix that specifies the value of the elements that are expected to be zero in the rotated pattern matrix (Browne, 1972a, 1972b). This target matrix guides the rotation of the factor loading matrix without imposing the traditional and less realistic principle of factorial simplicity (Kaiser, 1974).

Limitations: FACTOR offers more than 25 different rotations, some novel or less well-known. Which one should we choose? FACTOR does not have the typical user's manual. The applied researcher would appreciate some recommendations about when to use one rotation method or another.

FACTOR is a program specifically developed to meet the needs of novice, average, and advanced researchers in matters related to EFA. In this brief summary, we have only presented an overview of the most basic aspects, leaving the interested reader with options for performing second-order factor analysis, or for trying the intermediate factorial solutions between exploratory and confirmatory factor analysis. It should also be pointed out that the 9.2 version of FACTOR (Lorenzo-Seva & Ferrando, 2013) makes it possible to evaluate other non-linear approaches through Item Response Theory (IRT).

In summary, FACTOR allows many suitable options or combinations, always based on the fact that it automatically offers 1) criteria to assess the adequacy of the input correlation matrix for factor analysis and 2) criteria to assess the multivariate and univariate normality of the items' distributions. In addition, FACTOR is a program that is constantly being improved. The version in progress, 10.4.01, will include the following novelties: 1) factor estimation by means of the Diagonally Weighted Least Squares (DWLS) estimation method; 2) robust estimation for both ULS and DWLS; 3) revision of ML estimation methods to avoid convergence problems; 4) Bayesian estimation of the tetrachoric and polychoric correlation matrices; and 5) the possibility to obtain standard errors and confidence intervals for all the estimated parameters using bootstrapping procedures.

PRELIS (PRE-processor of LISREL)

With Prelis, the pre-processor of LISREL, it is possible to perform an Ordinal Factor Analysis, which is a non-linear analysis that uses full information estimation methods. Instead of analyzing polychoric correlations (where univariate and bivariate frequencies are used to estimate parameters), full information approaches use the whole response pattern to obtain information about parameter estimates (Bock, Gibbons, & Muraki, 1988). However, in this article, we only focus on the application of classic EFA, the topic of the review in Lloret et al. (2014). In this review, version 9.10 of LISREL was used, but the main differences from previous versions of LISREL will be pointed out.

1) Factorial model. Type of data and association matrix

Possibilities: When data are defined as continuous, PRELIS bases the analyses on Pearson correlations. When data are defined as ordinal, PRELIS bases the analyses on the polychoric or tetrachoric correlation matrices, depending on whether the data are polytomous or dichotomous, respectively. The corresponding variance-covariance matrices can also be analyzed. This is the option that appears by default. However, if the data are defined as ordinal and the MINRES (MINimum RESiduals) estimation method is used, which is equivalent to ULS (see the next section), then the polychoric (or tetrachoric) correlation matrix is analyzed, regardless of whether the covariance or correlation matrices are marked in the options (to use the variance-covariance matrix, see Jöreskog, 2002).

If the data are defined as continuous, the program offers information about the means and standard deviations of the items, as well as the skewness and kurtosis. In addition, it performs univariate normality contrasts automatically, and it offers the option of performing Mardia's multivariate normality test (Mardia, 1970). In the case of using ML, when there is severe multicollinearity for any of the variables, the program indicates this.

If the data are defined as ordinal, the program offers thresholds between response categories, as well as tests of the adequacy of the estimation of each bivariate correlation, in order to evaluate whether the latent variables underlying the ordinal items follow a normal distribution.

Thus, PRELIS offers the researcher a preliminary and automatic analysis of the distribution of the items.

Limitations: PRELIS does not offer information about the matrix's adequacy for factorization by means of the KMO measure or Bartlett's sphericity test.

2) Factor estimation methods

Possibilities: In addition to principal components (PC), which is not a true factorization method, PRELIS has two estimation methods: ML and MINRES. This latter method, proposed by Harman (1980), is equivalent to unweighted least squares (ULS), except for an orthogonal transformation of the factor loadings (Jöreskog, 2003). However, it should be mentioned that in version 9.10 of LISREL, the PC analysis does not appear as an estimation method within EFA, as in earlier versions, but rather it appears as a differentiated analysis.

Limitations: as in FACTOR, it does not have robust ML estimation methods or WLS.

3) Factor selection method

Possibilities it offers: In earlier versions of LISREL, the specific factor selection method depends on the estimation method used. When ML was used, the program fitted models with different numbers of factors (0, 1, 2, etc.) and offered goodness of fit indexes for the different models, specifically, the chi-square test and the RMSEA (Root Mean Square Error of Approximation) index for each model, and the significance of the chi-square difference between consecutive models. The stopping criterion in earlier versions of LISREL (8.30 and 8.80) takes into account various possible contingencies. If the null hypothesis that the model with zero factors fits the data cannot be rejected, the program stops. It would not make sense to continue with the analyses because the variables analyzed would be linearly independent. If the zero-factor model does not fit the data, the program increases the number of factors by one. If the probability associated with the chi-square statistic for a model with a *k* number of factors is greater than .10, the program considers that this model adequately represents the data, and it shows the solution for this model. If the probability associated with a model is less than .10, but the chi-square difference between this model and another model with an additional factor is greater than .10, the program stops because it considers that the difference between these models is not large enough to extract another factor, printing the solution for the most parsimonious model (i.e. with fewer factors). Finally, if the RMSEA value for a certain model with *k* factors is less than .05, the program prints the model solution with this number of factors. This latter criterion, according to Jöreskog, Sörbom, Du Toit and Du Doit (1999), is intended to guarantee that the number of factors will not be overestimated in large samples (Browne & Cudeck, 1993). If none of the aforementioned conditions are met, the program increases *k* (number of factors) by 1, and it continues the process until some of these conditions are met.

However, for LISREL 9.10, the criterion followed to maintain a certain number of *k* factors, regardless of the estimation method used, is the Kaiser criterion. Thus, the number of factors maintained will be equal to the number of eigenvalues greater than 1. Nevertheless, when using maximum likelihood (ML) with continuous data, the program continues to print the same decision table printed in previous versions, with the goodness-of-fit indices of the different models (and their comparison), from 0 factors to the number suggested by the Kaiser criterion. This information can be evaluated by the user to determine whether a lower number of factors could be sufficient to satisfactorily represent the data. In theory, it is also possible to set the number of theoretically expected factors. However, if the number of factors requested is different from the number extracted following the Kaiser criterion, the program does not print a solution.

Limitations: PRELIS does not offer additional criteria, such as PA or MAP, a variety of goodness of fit indexes, or the scree-test (although this can be obtained from the eigenvalues).

4) Factor rotation methods and item assignment

Possibilities: With both estimation methods (ML and MINRES), the following solutions are offered: 1) Non-rotated, 2) with orthogonal rotation, specifically VARIMAX (Kaiser, 1958), 3) with oblique rotation, specifically PROMAX (Hendrickson & White, 1964), and 4) the reference variable solution, which also offers the correlations between factors. This latter solution is obtained using the TSLS (Two-Stage Least Squares; see Jöreskog et al., 1999) estimation method, and the reference items chosen are the items in the PROMAX solution with the largest factor loading in the corresponding factor. The advantage of using this latter method is that it provides standard errors and *t* values for the factor loadings, except for the reference items. Consequently, it is possible to determine whether the parameter estimates are statistically significant. If one wants to estimate the reference variable solution using ML, it is possible to employ specific commands (see syntax examples in the manual). In this latter case, it is advisable to use the variance-covariance matrix as input, in order to obtain correct standard errors (see, Jöreskog et al., 1999).

In summary, PRELIS allows the use of different input matrix depending on the nature of the data and the most adequate model, but it is completely limited in terms of the factor selection method, and it has few estimation (only ML and MINRES) and rotation (VARIMAX as orthogonal and PROMAX as oblique) methods.

MPlus

1) Factorial model. Type of data and association matrix

Possibilities: MPlus offers different possibilities, allowing the researcher to analyze the appropriate matrix according to the nature of the data. When nothing is indicated in the input file, the program assumes that the data are continuous; when they are ordinal or categorical, this must be indicated with the corresponding instruction. For continuous items, it will estimate the Pearson correlation matrix; for ordinal polytomous items, it will estimate the polychoric correlation matrix; and for dichotomous data, it will estimate the tetrachoric correlation matrix. If the researcher thinks his/her data can be modeled more robustly and simply by using the linear model, then he/she will have to define them as continuous, as occurs with PRELIS.

In the case of categorical items, the program offers preliminary information about the distribution of the subjects in the different item response categories. Specifically, it offers the percentage and number of subjects who answered each of the possible alternatives.

Limitations: MPlus does not offer adequacy tests (such as KMO, for example) to evaluate the adequacy of the correlation matrix for its factorization. Nor does it offer preliminary and automatic tests to evaluate the fit of the data to normality.

2) Factor estimation methods

Possibilities it offers: MPlus offers a wide variety of factor estimation methods, and it allows the researcher to choose some of the recommended ones.

When the items have been defined as continuous, MPlus offers 4 possible estimation methods: ML, robust ML (MLM, MLMV), and ULS. With regard to the robust ML estimation methods, they offer robust estimations of the standard errors and the chi-square test. The MLM option provides a mean-adjusted chi-square model test statistic, so that the Satorra-Bentler chi-square is offered (Satorra & Bentler, 2001). The MLMV option produces a mean and variance adjusted chi-square test of model fit.

When there is at least one categorical item, MPlus also offers 4 possible estimation methods: weighted least squares (WLS), robust WLS (WLSM, WLSMV), and ULS. The default option is WLSMV; and the least recommendable option is ULS, which is less stable and, when combined with the tetrachoric correlation matrix, does not offer goodness of fit indexes (which also occurs with FACTOR). As in the case of the robust ML estimators, the robust WLS estimators offer robust estimations of standard errors and the chi-square statistic, and the name of each estimation method refers to whether the solution provides a mean-adjusted chi-square (WLSM) or a mean and variance adjusted chi-square (WLSMV).

Therefore, we can see that Mplus offers robust estimation methods for the violation of the multivariate normality assumption, such as WLS, Robust WLS (WLSM, WLSMV), or Robust ML (MLM, MLMV). Choosing one or another will depend on what type of matrix is more appropriate as input, considering the types of items.

3) Factor selection method

Possibilities: MPlus offers methods based on residual minimization. Specifically, it offers RMSEA and SRMR (Standardized Root Mean Square Residual).

This program requires the researcher to set the number of expected factors, and the output offers the result of the different models tested: from a one-factor solution, to a solution with the specified number of factors, up to a maximum of 9. Therefore, it makes it possible to compare the fit of alternative models based on their goodness-of-fit indexes. This comparison can be carried out based on incremental fit indexes (ΔRMSEA, ΔCFI and ΔNNFI), following the criteria recommended in the literature (e.g. Chen, 2007; Cheung & Rensvold, 2002; Widaman, 1985).

Limitations: It does not offer some of the "extra" criteria recommended, such as PA or MAP. It only employs the model comparison strategy.

4) Factor rotation methods and item assignment

Possibilities it offers: MPlus offers a wide variety of orthogonal and oblique rotation methods: VARIMAX, PROMAX, QUARTIMIN, OBLIMIN, GEOMIN, CF-VARIMAX, CF-QUARTIMAX, CRAWFER, CF-EQUAMAX, CF-PARSIMAX, CF-FACPARSIM and TARGET. All of the rotation methods are available with both orthogonal and oblique rotation, except VARIMAX, which is orthogonal, and PROMAX and QUARTIMIN, which are oblique. The default rotation method is oblique GEOMIN.

As presented in recent studies (Ferrando & Lorenzo-Seva, 2014), the rotation method used somewhat determines the more exploratory or confirmatory nature of the analysis performed. Specifically, the Target rotation available in MPlus allows a more confirmatory approach under the label of EFA. This rotation makes it possible to propose a target matrix that specifies the value of some of its elements, specifically those expected to be zero in the rotated pattern matrix. Therefore, this target matrix guides the rotation of the factor loading matrix. A more detailed description of the criteria followed when this rotation method is applied is provided in the study by Ferrando and Lorenzo-Seva (2014) under the name "partially-specified procrustean rotation", and it can also be consulted in the study by Browne (2001) under the name "rotation to a partially-specified target matrix". In addition, the items' assignment to the factors is left to the researcher, who will apply the recommended criteria to determine which items belong to each factor.

Limitations: In MPlus, as in FACTOR, the applied researcher can feel overwhelmed by the large number of rotation options the program offers, and he/she may need some recommendations about when to use one rotation method or another.

In summary, MPlus can be used efficiently because it offers a wide range of possibilities, although it also has some limitations, mainly related to the factor selection criteria.