anales de psicología, 2014, vol. 30, nº 3 (octubre), 1151-1169
http://dx.doi.org/10.6018/analesps.30.3.199361

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

Exploratory Item Factor Analysis: A practical guide revised and updated

Susana Lloret-Segura*, Adoración Ferreres-Traver, Ana Hernández-Baeza, and Inés Tomás-Marco

Unidad de Aplicaciones y Desarrollos Psicométricos.
IDOCAL. Instituto de Investigación en Psicología de los Recursos Humanos, del Desarrollo organizacional y de la Calidad de Vida Laboral
Facultad de Psicología. Universitat de València (Spain)

Título: El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada.
Resumen: El Análisis Factorial Exploratorio es una de las técnicas más usadas en el desarrollo, validación y adaptación de instrumentos de medida psicológicos. Su uso se extendió durante los años 60 y ha ido creciendo de forma exponencial al ritmo que el avance de la informática ha permitido. Los criterios empleados en su uso, como es natural, también han evolucionado. Pero los investigadores interesados en asuntos sustantivos que utilizan rutinariamente esta técnica permanecen en muchos casos ignorantes de todo ello. En las últimas décadas numerosos trabajos han denunciado esta situación. La necesidad de actualizar los criterios clásicos para incorporar aquellos más adecuados es una necesidad urgente para hacer investigación de calidad. En este trabajo se revisan los criterios clásicos y, según el caso, se sustituyen o se complementan con otros más actuales. El objetivo es ofrecer al investigador aplicado interesado una guía actualizada acerca de cómo realizar un Análisis Factorial Exploratorio consonante con la psicometría post-Little Jiffy. Esta revisión y la guía con las recomendaciones correspondientes se han articulado en cuatro grandes bloques: 1) el tipo de datos y la matriz de asociación, 2) el método de estimación de factores, 3) el número de factores a retener, y 4) el método de rotación y asignación de ítems. Al final del artículo hemos incluido una versión breve de la guía.
Palabras clave: AFE; ACP; matriz de asociación; estimación de factores; número de factores; rotación de factores.

Abstract: Exploratory Factor analysis is one of the techniques used in the development, validation and adaptation of psychological measurement instruments. Its use spread during the 1960s and has been growing exponentially thanks to the advancement of information technology. The criteria used, of course, have also evolved. But the applied researchers, who use this technique as a routine, remain often ignorant of all this. In the last few decades numerous studies have denounced this situation. There is an urgent need to update the classic criteria. The incorporation of the most suitable criteria will improve the quality of our research. In this work we review the classic criteria and, depending on the case, we also propose current criteria to replace or complement the former. Our objective is to offer the interested applied researcher updated guidance on how to perform an Exploratory Item Factor Analysis, according to the "post-Little Jiffy" psychometrics. This review and the guide with the corresponding recommendations have been articulated in four large blocks: 1) the data type and the matrix of association, 2) the method of factor estimation, 3) the number of factors to be retained, and 4) the method of rotation and allocation of items. An abridged version of the complete guide is provided at the end of the article.
Key words: AFE; PCA; matrix of association; factors' estimation; number of factors; factors' rotation.

Introduction

Exploratory Item Factor Analysis (EFA) is one of the most commonly applied techniques in studies related to test development and validation, as it is the technique par excellence used in exploring the set of latent variables or common factors that explain the answers to the items of a test. If there is a query about material published in scientific journals, regarding this type of analysis two clearly separated tendencies are found. Firstly, the most numerous, where the technique is applied to identify the structure underlying test items. This is an instrumental use of the analysis technique. The other, less numerous, where we study and compare the different criteria usually applied in carrying out EFA, with new or not so new but less popular criteria. These studies investigate the most appropriate decisions, according to the conditions where the technique is applied. In this case, the decision regarding the technique itself is the object of study. This attention is deserved as the phases the factor analysis passes through require the application of decision criteria that, like most things, have improved over time. Despite these changes, traditional criteria, some quite outdated, still coexist today alongside the most modern ones, and that is where the maze metaphor comes in: the researcher, sometimes taken aback by the amount of possibilities or in other cases being unaware of new possibilities, carries on, trusting their own and others' experience or a combination of both, but unsure of having found the main exit or the service exit.

The paradox in this context is as follows: methodologists seek and find better criteria for applying EFA, while unanimously warning that the classics are often inadequate and dangerous. This opens new and better ways out of the maze, and states that some paths are misleading. On the other hand, the researchers who make instrumental use of the EFA are seemingly unaware that the new paths are better and that the old ones are dangerous, so they continue through the maze, oblivious to the signs indicating where to go.

In our view, there is a clear failure in communication, or signaling as regards the maze. Maximum diffusion must be given to these new criteria, as nowadays it is known when and why they work better than the previous ones. On the other hand, we must also demand that they be applied in current research, as we know they perform better than the classic versions. As pointed out earlier, not just any path leads out of the maze.

The aim of this study is to contribute to the dissemination and application of these new criteria or standards (and some not so new, which are apparently systematically ignored).This will be done in three ways: first, by offering a revision and update of the classic standards, and by presenting those that replace them. Second, by reviewing and summarizing the extent to which a variety of software available

*** Dirección para correspondencia [Correspondence address]:**
Susana Lloret Segura. Department of Methodology and Behavioral Sciences. Faculty of Psychology. University of Valencia. Blasco Ibáñez, 21. 46022 Valencia (Spain). E-mail: Susana.Lloret@uv.es

on the market - SPSS, FACTOR (Lorenzo-Seva & Ferrando, 2006, 2012b, 2013), LISREL (Jöreskog & Sörbom, 2007), and MPlus (Muthén & Muthén, 2007) - allows or limits the application of these new standards. And third, by using that software to analyze the factor structure of three sets of scales-D-48 (Anstey, 1959. Adapted to Spanish by the R & D Department of TEA Editions, 1996), Self-esteem and Self-concept, and Strength and Flexibility (Marsh, Richards, Johnson, Roche, & Tremayne, 1994), using real data for each set with the particularity that the scale data fits inadequately, moderately or satisfactorily, respectively,-to the assumptions of the classic EFA model. These three practical cases with real data will allow us to compare the consequences of choosing a type of sotfware, as well as the consequences of choosing a specific option from a number of available choices within the same software, when the data are challenging. Our goal is clear: to illustrate how the appropriate or inadequate application of EFA can lead to very different conclusions.

This goal is not new, in the last two decades at least 6 different studies reviewing the use of EFA in empirical research in Psychology have highlighted the limitations characterizing the empirical studies published. These are summarized below.

Ford, MacCallum, and Tait (1986) reviewed articles published in three journals (*Journal of Applied Psychology*, *Personnel Psychology*, *Organizational Behavior and Human Performance*) in the period between 1975 and 1984 reviewing a total of 152 studies; Ford et al. (1986) concluded that the application of EFA in the reviewed studies was deficient. Almost 10 years later, Hinkin (1995) reviewed the use of EFA in the development of 277 scales. These appeared in 75 articles published in relevant academic journals from 1989 to 1994, and came to the same conclusion. Fabrigar, Wegener, MacCallum, and Strahan (1999) reviewed 217 articles published in two psychology journals (J*ournal of Personality and Social Psychology*, *Journal of Applied Psychology*) in the period 1991-1995. In this paper, a similar conclusion to Ford et al. (1986) was reached, although there was a tendency was to follow some newer recommendations. More recently, Conway and Huffcutt (2003) conducted a review of the use of EFA in the same three organizational psychology journals reviewed by Ford et al. (1986) (*Journal of Applied Psychology*, *Personnel Psychology*, *Organizational Behavior and Human Performance*) in the period 1985-1999 (a total of 371 studies were reviewed). Conway and Huffcut (2003) again confirmed as light tendency to use more modern criteria; although this was still uncommon. Finally, Henson and Roberts (2006) reviewed 60 articles published in four journals (*Educational and Psychological Measurement*, *Journal of Educational Psychology*, *Personality and Individual Differences*, and *Psychological Assessment*) published until 1999. Their conclusion was that the application of EFA in the reviewed studies showed a pattern of common errors to be avoided.

Izquierdo, Olea, and Abad (2013) recently presented a paper analyzing the use of EFA in 2011 and 2012 in the three Spanish journals with the greatest impact factor in the last 5 years: *International Journal of Health and Clinical Psychology*, *Psicothema*, and *Spanish Journal of Psychology*. Their conclusions were similar to the previous: high rates of incorrect or unjustified decisions.

We will now analyze erroneous decisions in detail, usually based on classic or outdated criteria or standards, and what other recommendations or guidelines can be used instead.

Determination of adequacy of the Exploratory Factor Analysis

When we speak about tests in Psychology we often refer to latent variables or traits causing responses to in that test. The aim of the test is to evaluate to what extent a person is characterized by a particular trait or latent variable, known as extraversion, "g" factor, or stress, based on the observed responses to a particular and well-chosen set of items - observed variables. (If the test is multidimensional, we then have multiple traits or latent variables). The EFA, Principal Component Analysis (PCA) and Confirmatory Factor Analysis (CFA) are techniques used for this purpose. However, they are not interchangeable. When should we choose either? Let us see in detail.

When do we apply Factor Analysis and when Principal Components?

The key lies in the exact aim of the analysis. If it is to identify the number and composition of the common factors (latent variables) needed to explain the common variance of the analyzed set of items, then applying EFA is appropriate. In this case the algebraic representation of the model for $m \leq p$ common factors has as equation:

$$X_1 = v_{1(1)} F_{(1)} + v_{1(2)} F_{(2)} + \dots\dots\dots + v_{1(m)} F_{(m)} + e_1$$
$$X_2 = v_{2(1)} F_{(1)} + v_{2(2)} F_{(2)} + \dots\dots\dots + v_{2(m)} F_{(m)} + e_2$$
$$\vdots$$
$$X_p = v_{p\,(1)} F_{(1)} + v_{p\,(2)} F_{(2)} + \dots + v_{p\,(m)} F_{(m)} + e_p$$

where X_j, F_i, and e_j refer to a person's score in item X_j, the common factor F_j, and the specific factor e_j, m: number of common factors, p: number of items, F: common factor, $v_{j(i)}$ Weight for the n^{th} common factor associated with the i^{th} observed variable or item, i = 1, 2, .., m; j = 1, 2, ..., p; e_j: uniqueness; r, j = 1, 2, ..., p.

On the other hand, if the aim is to identify the number and composition of components needed to summarize the scores obtained on a large set of observed variables, then applying PCA is appropriate .This method explains the maximum percentage of variance observed in each item from a lower number of components summarizing this information. In this case, the algebraic representation of the model for $m \leq p$ major components has as equation:

$$PC_1 = w_{(1)1} X_1 + w_{(1)2} X_2 + \ldots + w_{(1)p} X_p$$
$$PC_2 = w_{(2)1} X_1 + w_{(2)2} X_2 + \ldots + w_{(2)p} X_p$$
$$..$$
$$PC_m = w_{(m)1} X_1 + w_{(m)2} X_2 + \ldots\ldots\ldots + w_{(m)p} X_p$$

where X_j and PC_j refer to the person's score in item X_j, and the component PC_i, m: number of principal components, p: number of items or observed variables, x: items or observed variables, PC: main components, $w_{j(i)}$ weight chosen for the observed variable j^{th} to maximize the ratio of the CP variance (i) to the total variance, i=1, 2,.., m; j= 1, 2,…, p.

The differences between the first option (EFA) and the second (PCA) are clear. The observed variables (items) are the independent variables in PCA, but dependent on EFA. At conceptual (and formal) level this difference is substantial. However, for numerous reasons there has been a great amount of confusion over the decades, so that one technique, PCA, has been systematically applied to achieve the aim of the other, EFA. We will see the reasons why.

Firstly, both techniques often appear together as data reduction techniques in the most common statistical packages. This makes them appear interchangeable; but they are actually quite different. The PCA is consistent with the conceptualization of formative measures, where indicators are thought to be the cause of a possible construct (Bollen & Lennox, 1991; Borsboom, Mellenbergh, & van Heerden, 2003; Joliffe, 2002), although the model itself makes no explicit reference to latent variables. In fact, components are neither "latent" variables, nor are items any indirect "measure" of them. The components are "composites" of the observed variables that fulfill the mission of reproducing the maximum variance of each observed variable with the minimum number of composites. The focus is, therefore, on the diagonal of the matrix, in the variances of the observed variables. The condition that the components are not correlated with each other is part of the procedure in achieving this aim.

The EFA assumes that the observed variables are indicators of a number of common latent factors or variables. If we analyze a set of items chosen to measure a single factor, each observed variable or item analyzed is carefully selected to reflect some feature of the factor to be measured along with it. The essential idea is that people with different levels in the common factor will provide different answers to that item, precisely because the factor provokes different responses (the item is a manifestation of that factor). The independent variable is the factor which produces different answers in items. The items are the dependent variables in this design. In other words, under EFA the measures are assumed to be "reflective" or manifestations of the underlying constructs (Bollen & Lennox, 1991; Edwards, 2011; Harman, 1976; Kim & Mueller, 1978). Furthermore, however careful the selection of items, they cannot be perfect indicators of the corresponding common factor. A part of the item´s variability will be directly produced by the factor measured by the item, but another part will not. Under the classical model or classical test theory, we can estimate the part of the variance of each item explained by the common factor underlying that set of items, precisely from the common variance shared between that item and the rest of items that measure that same factor (this part of the variance of the item is called communality). The remaining variance of the item is non-common variance (uniqueness) which does not contribute to the measurement of the common factors, and therefore is not included in the process of identification and estimation of those factors. Herein lies the second great difference between EFA and PCA: the inclusion of an error term (e). This error term does not exist in the PCA. The PCA does not distinguish between common variance and non-common variance (see Joliffe, 2002, Chapter 7, for a detailed and accessible treatment of similarities and differences between both models for data reduction).

Finally, it should be pointed out that the common factors are not aimed at explaining the maximum amount of variance of each item (as with components), *only the common variance of each item with the rest*, and that common variance is no longer in the elements of the diagonal matrix, *but in the elements outside the diagonal, which are expressed in terms of covariance or correlations. These are the elements that the researcher will try to explain.*

What is the classic recommendation?

Clearly, when we analyze the items of a test seeking its corresponding factor structure we are in the second of the above scenarios, i.e. under the model of factor analysis. However, for some decades, the computational difficulty of the procedures aimed toward the estimation of common factors (CF) characteristic of EFA far exceeded that of the identification of principal components (PCA) to the point where it would not be possible to obtain a solution by means of EFAs (or the solution would be inadequate, with Heywood cases, i.e., items with commonalities of 1 or even greater, or non-positive definite matrices). The use of PCA as a method of finding an initial solution identifying the number of dimensions needed to explain the set of items did have these problems, and that solution could be rotated in search of the most adequate (orthogonal or oblique) structure, in the same way as an initial solution obtained by a CF procedure would have been rotated. Therefore, the PCA became the simplest and most effective method of estimating / extracting the underlying "factors" (actually components). Some studies suggested it was reasonable to apply PCA in the context of factor estimation when 1) the number of items per factor was high, and 2) the items contained little measurement error, since in these particular conditions the solution obtained by either procedure was virtually equivalent (Thompson, 2004; Velicer & Jackson, 1990). Thus, the idea of equivalence between both techniques spread. Once this equivalence was shown for a set of very particular conditions, researchers began intensively using PCA as a way of extracting "factors" in the context of EFA, regard less of whether these particular conditions were given or not . There were few alternatives, so the tacit agreement to use PCA as if it were a CF analysis technique became universal and therefore the classic recommendation.

What is the current recommendation?

The computational efficiency of PCA compared to EFA does not exist anymore. There are new factor estimation options (such as the ULS factorization method that we will see later in the paper) enabling us to apply EFA in conditions once previously impossible. So the use of PCA to estimate factors in the context of EFA no longer makes sense. In contrast, there are empirical and simulation studies strongly discouraging the use of PCA as if it were EFA, as both types of analyses can lead to quite different solutions (Ferrando & Anguiano-Carrasco, 2010; Gorsuch, 1997a; Vigil-Colet et al. 2009).

When applied in situations where a factor model is more adequate, the use of PC as a factor estimation method means ignoring the measurement error, which spuriously increases factor loadings and variance percentages explained by factors, and can produce an overestimation of the dimensionality of the set of items (see Abad, Olea, Ponsoda, & García, 2011, Ferrando & Anguiano-Carrasco, 2010, Gorsuch, 1997a). This occurs when seeking components that explain the total variance (common variance plus error variance, jointly considered) instead of only accounting for commonality. Either way, the interpretation of the solution obtained through PCA could be erroneous.

Why is PCA still in use?

The message has not been received by everyone and there is also a trend toward ambiguity: factor analysis programs include the PCA method among their options of factor estimation methods, and even as the default option (and therefore as reference). Unfortunately this trend has not diminished.

Ford et al. (1986) covered the period 1975-1984 finding that 42% of the reviewed studies used PCA, 34% EFA, and the rest did not report the model used. Hinkin (1995) analyzed the period 1989-1994 and found that 33% of the reviewed studies used PCA, whereas (only 20% used the EFA model in any variant). Fabrigar et al. (1999) found that this percentage increased to 48%.

Conway and Huffcutt (2003) recently performed a review of the use of EFA in the same three organizational psychology journals reviewed by Ford et al. (1986) and reported that this model remained the most common, in 39.6% of cases. Finally Henson and Roberts (2006) offered a greater percentage of PCA use in their study: 56.7%. In Spain, Izquierdo et al (2013) presented the same trend as Henson and Roberts. These results clearly reveal that it must be stressed that EFA must be applied if the aim is to analyze correspondence between a series of items and the set of factors to measure those items.

When do we apply Exploratory Factor Analysis and when Confirmatory?

EFA does not allow the researcher to define which items measure which factors, nor the pattern of relationships assumed between the factors themselves, apart from whether they are all related to each other or not. It is called exploratory as we can only determine the number of factors we expect, but neither the composition nor specific pattern of relationships among factors. In contrast, the CFA is characterized by allowing the researcher to define how many factors are expected, which factors are related to each other, and which items are related to each factor.

What is the classic recommendation?

The classic recommendation (e.g., Mulaik, 1972), which is still valid (e.g., Matsunaga, 2010), differentiates between exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) depending on its aim. From this perspective, both methods are used to evaluate the factor structure underlying a correlation matrix, but whereas EFA is used to “build” the theory, CFA is used to "confirm" it. Thus EFA is used when the researcher knows little about the variable or construct of interest and this approach helps to identify the latent factors underlying the manifest variables, in addition to the patterns of relations between latent and manifest variables. On the other hand, when we have a clear idea about the variables under study, the use of CFA allows testing the hypothesized structure and checking if the hypothesized model adequately fits the data.

What is the current recommendation?

There are currently two trends. The first, derived directly from the classic approach, recommends making sequential use of both types of analysis, as long as the sample size allows it. This means randomly dividing the sample into two subsamples and exploring the factor structure underlying the items in the first sample (with an exploratory factor analysis), then trying to confirm that structure in the other half of the sample, by means of a confirmatory factor analysis (Anderson & Gerbing, 1988; Brown, 2006).

From the second trend, the distinction between EFA and CFA is questioned regarding its (exploratory / confirmatory) purpose, indicating that such differentiation is unclear and brings a series of problems (e.g., Ferrando & Anguiano-Carrasco; 2010). To begin with, it is easy to understand that most psychometric applications of factor analysis are somewhere between the total absence of information on the variables under study and the clear definition of their factor structure. For this reason, Ferrando & Anguiano-Carrasco (2010) proposed establishing the differentiation between both approaches, not according to their aim, but in the restrictions imposed. Thus instead of considering EFA and CFA as two qualitatively distinct categories, they should be considered as two ends of a continuum. Thus, EFA (nonrestrictive) imposes minimum restrictions to obtaining an initial factor solution, which can be transformed by applying different rotation criteria. And the CFA (restrictive) imposes much stronger constraints that allow a single solution to be tested, the fit of which can be evaluated using different goodness of fit indices (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004).

Continuing with this second trend, various authors have shown that CFA fails to confirm factor structures clearly supported by corresponding exploratory analyses as it is too restrictive(e.g. Ferrando & Anguiano-Carrasco, 2010; Ferrando & Lorenzo-Seva, 2000; Marsh et al., 2009, 2010; Marsh, Liem, Martin, Morin, & Nagengast, 2011; Marsh, Morin, Parker, & Kaur, 2014). In the CFA model, some hypotheses posit that some item factor loadings are zero (specifically the loadings on the factors items do not intend to measure). In many cases this restriction is not realistic, especially if factors are correlated. This problem has been thoroughly discussed by Ferrando & Lorenzo-Seva (2000), and as the authors indicate causes an accumulation of specification errors. Most items analyzed do not act as markers in practice (i.e., they are not factorially simple), and therefore, they present smaller, but nonzero, cross-loadings on the other factors that they do not intend to measure. When these crossloadings are forced to be zero, as in CFA, the goodness-of-fit of the model deteriorates, as the residuals accumulate with each specification error. Bad-of-fit is greater the longer the test (more errors accumulated) and the larger the sample (the greater the power).

To solve this problem, the current recommendation presents different alternatives. One is ESEM (Exploratory Structural Equation Modeling), representing a hybrid between EFA and CFA. ESEM is a semi-confirmatory alternative, since it is midway between both analyses strategies in the restriction continuum indicated above, integrating the advantages of both approaches.

The ESEM (e.g., Marsh et al., 2011; Morin, Marsh, & Nagengast, 2013) is a similar approach to CFA, but item factor loadings on factors different from those the items intend to measure (or crossloadings) are not fixed to zero. Thus, it is less restrictive than CFA, and fits better to the reality of the measures used in Psychology (where items are not normally perfect "markers" or "indicators" of the construct they measure, but it is common for the items to be factorially complex and have lower but not zero weights in the other factors). Therefore, the ESEM obtains better fit than the CFA. We may say that this approach moves away from the "simple structure" proposed by Thurstone (1947) for easy interpretation of EFA, which has long been the norm, and is closer to the current trend proposing models that approach well enough to reality, but without trying to reproduce it, as this is either impossible, or implausible.

Another alternative option, initially made by Mulaik (1972) is to specify a directly interpretable solution using one or two markers per factor. This alternative has also been recently considered by McDonald (1999, 2000, 2005) in the "independent-cluster basis" (ICB) solution. The ICB concept refers to the factor solution of a multidimensional test, where each factor is defined by a small number of factorially simple items (markers). Specifically, the requirement is there are at least 3 markers per factor if - factors are not correlated and at least 2 markers per factor if factors are correlated (McDonald, 1999). In a semi-restricted solution, the remaining items may be factorially complex. As indicated by Ferrando & Lorenzo-Seva (2013), compliance with the ICB condition is enough to identify a solution without rotational indeterminacies, a disadvantageous regarding interpretation. Finally, another alternative is to use a semi-specified or Procrustean rotation to a target factor pattern matrix which can be applied with the FACTOR program. A practical application of this alternative can be consulted in the work of Ferrando, Varea, and Lorenzo (1999).

There is therefore no universal recommendation as to when to apply EFA or CFA. However, regardless of whether either (or a sequence of both) is chosen, researchers must adequately justify their choice.

Matters related to design: selection of items, sample size and composition, and number of items per factor

As with any research, the utility and generalizability of - results obtained with EFA will depend on the adequacy of the research design, i.e. on the selection of the variables to be measured- the sampling procedure used, and the sample size- among other decisions. Empirical studies tend to neglect this phase of the research (Ferrando & Anguiano-Carrasco, 2010). We will now look at each of these aspects in detail.

Sample size and composition

How to select the most appropriate items to build the test

One of the first issues to be decided is the subset of items that will configure the initial version of the test. If the subset omits relevant aspects of the latent variable to be measured, there will be less common variance than required in corresponding analysis and resulting common factors will be weaker for being insufficiently defined. On the contrary, if irrelevant items are introduced, additional common factors may appear or it will be difficult for the intended common factors being measured to emerge. Therefore adequate choice of items plays a decisive role in the clarity of the identified factor structure.

Classic recommendation

The classic recommendation in the selection of items is to clearly and thoroughly define the construct to be measured and, from that definition, select items covering all the relevant aspects of the definition. From the classical viewpoint this refers to content validity. Using some empirical criteria when selecting items is also advisable, such as the corrected homogeneity index (item-total correlation without the item analyzed) and the alpha coefficient if the item is eliminated from the scale (or subscale, depending on whether the test is one-dimensional or made up of various subscales).

Current recommendation

The classic recommendations are still in use today. What is new is that additional recommendations have been incorporated that further refine the process of item selection. The accumulated experience reveals the importance of both substantive and methodological considerations, summarized below:

A) The use and abuse of redundant items may spoil the resulting factor structure. Redundant items expressing the same idea with slightly different wording have been traditionally used to evaluate the consistency of people's responses, but also to raise the internal consistency of the scales (Ferrando & Anguiano-Carrasco, 2010). The problem arises because these redundant items naturally share more variance than is directly explained by the common factor. These pairs or triples of redundant items also share part of the unique variance. And when this occurs, additional common factors emerge, which seem difficult to identify and explain, especially after rotating the initial solution.

B) Also playing a key role are item response distributions and the number of item response options. If the Pearson correlation or covariance matrix is analyzed (because the program used does not allow otherwise), the items must be continuous variables. Otherwise, as happens when the items to be analyzed are polytomous (e.g., Likert type), using items with at least five response alternatives and with approximately normal distributions allows us to adequately approach the assumption of continuity . Items with fewer response categories or with non-normal distributions should be analyzed according to their ordinal nature, i.e. using the polychoric correlation matrix (for polytomous items, or the tetrachoric correlation matrix, in the case of dichotomous items) (Bandalos & Finney, 2010). We will return to this point later.

How many items per factor should be included?

Guadagnoli and Velicer (1988; MacCallum, Widaman, Preacher, and Hong, 2001; MacCallum, Widaman, Zhang, and Hong, 1999) showed that the number of items per factor interacts with the size of the commonalities of items and with the sample size.

Common practice is to select at least three items per factor. However, this is counterproductive, as it compromises the stability of results (Velicer & Fava, 1998), especially when the sample size is below 150 (Costello & Osborne, 2005).

As a general rule, the more items accurately measuring a factor, the more determined the factor and the more stable the factor solution. The revised studies point to a minimum of 3 or 4 items per factor, only with a minimum of 200 cases (Fabrigar et al., 1999; Ferrando & Anguiano-Carrasco, 2010).

Suitability of sample

What minimum sample size is needed? What aspects of the sample composition should be considered?

What sample is needed to make a solution stable and generalizable? This is a complex problem. The question of sample size and composition in EFA has been the subject of research for decades. Convenience samples are commonly used (often university students), but there are two problems: lack of representativeness and attenuation and restriction of variance (Ferrando & Anguiano-Carrasco, 2010). Larger samples are obviously better than smaller, but the researcher cannot always access large sample sizes, which has led to a wide range of studies offering a large list of sometimes incorrect recommendations about the minimum sample size in EFA (Guadagnoli & Velicer, 1988; Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005).

Classic recommendation

Many of these recommendations on sample size have been the object of numerous empirical and simulation studies (Arrindell & van der Ende, 1985; MacCallum et al., 1999; Velicer & Fava, 1998), where two approaches are distinguished: 1) those suggesting a minimum size (N), and 2) those that defend a specific ratio of individuals per item (N/p).

Some classic studies within the first approach (Barrett &Kline, 1981; Guadagnoli & Velicer, 1988) suggest a minimum sample size (N), ranging from 50 to 400 subjects. Comrey and Lee (1992, p.217) suggested the adequacy of sample size could be assessed on the following scale: "50 -very poor; 100 -poor; 200-fair; 300-good; 500-very good, 1000 or more–excellent." So one of the classic recommendations par excellence is that a size of 200 cases or more (except in clinical samples) is sufficient for most descriptive and psychometric analyses of items, if the test to be validated is quite short. Although reaching 500 or more cases is recommended when possible (MacCallum et al., 1999).

On the other hand, from the second approach, based on the person / item ratio (N/p), the most common recommendation is the "rule of 10": a sample 10 times greater than the number of items (Velicer & Fava, 1998), and the ratio of 5: 1. In a classic work Gorsuch (1983) suggested a ratio of 5 subjects per variable, and a size not fewer than 100 people.

Current recommendation

At present, different simulation studies reveal that sample size is a factor that interacts with other aspects of the design and the nature of the data, such as the matrix serving as input to the EFA, the number of items that define the factor, the homogeneity of the sample and, in particular, item commonalities (Beavers et al., 2013; Fabrigar et al., 1999; Ferrando & Anguiano-Carrasco, 2010).

Regarding the input matrix, we distinguish between the product-moment correlation matrix and the polychoric correlation matrix. The former is calculated directly on empirical data, while polychoric correlations are obtained from indirect estimators in an iterative way and are generally much more unstable. Therefore, an EFA based on the polychoric correlation matrix will require larger samples than one based on product-moment correlations in order to reach the same level of precision and stability, if the remaining conditions remain constant.

Regarding other aspects mentioned above, recommendations are summarized as follows:

1) Optimum condition: When factor loadings are greater than .70, and the number of variables per factor is adequate - at least 6 items -, a sample size of 150 or 200 cases appears sufficient to obtain accurate estimates of the EFA coefficients (MacCallum et al., 1999, Preacher & MacCallum, 2003). There is even evidence that 100 cases will suffice
2) When there are three factors with three or four items each, or when there are more items and factors but commonalities above .80 (Bandalos & Finney, 2010; Costello &Osborne, 2005; Guadagnoli & Velicer, 1988)
3) Moderate condition: When commonalities are between .40 and .70, and the number of variables per factor is 3-4 items, a size of 200 cases is also accepted, and finally,
4) Minimum condition: When commonalities are low, around .30, and the number of variables per factor is 3 items, a minimum sample of 400 cases is required (Conway & Huffcutt, 2003); Even 500 or more would be needed to obtain sufficiently accurate estimates (Hogarty et al., 2005).

Nowadays, the classic N/p criteria and traditional "recipes" such as 10 times more subjects than items, among others, are completely discouraged, since they have no solid basis (Bandalos & Finney, 2010; Ferrando & Anguiano, 2010) . There is actually no obvious recipe, as the recommended minimum size depends on all factors listed. Logically, the larger the sample size the more we believe the solution obtained is stable, especially if commonalities are low, or when there are many possible factors to extract and/or few items per factor. However, one thing seems clear: to evaluate the quality of a test, a sample size of at least 200 cases is recommended, even in optimal conditions of high commonalities and well-determined factors (Ferrando & Anguiano-Carrasco, 2010).

Types of data and association matrices. Adequacy of data to Factor Analysis

Classic FA has been developed on the assumption that the items are linearly related to the factors they measure, and that their relationships are also linear (there are variants of non-linear FA, but these are not the object of this work). If variables are continuous or close enough to this condition, Pearson's product-moment correlation matrix, or the variance-covariance matrix, will adequately summarize the relationships between the items (the former more frequent than the latter, especially in EFA, see Brown (2006), pp. 40-42). However, in Psychology it is common to use the Pearson correlation matrix or the variance-covariance matrix obtained on ordinal items without determining if their distributions meet this assumption. Some programs do not allow the use of another type of matrix.

Classic recommendation

The classic recommendation is to use Pearson's correlation matrix as input matrix. This type of correlation is adequate in determining the linear relationship between two continuous variables that are preferably normally distributed. Items are not continuous variables, but ordinal and discret. However, when their distributions are approximately normal, Pearson's correlation coefficient is still a good method of estimating the two-item relationship. And when is that distribution roughly normal? In the case of dichotomous items when the difficulty indices are moderate and homogeneous (between .40 and .60). In other words when the items present intermediate difficulties, and consequently "symmetric" distributions. The items with asymmetric distributions are especially problematic if they appear in both directions, as they give rise to nonlinear relations. This problem, typical of tests with very easy and very difficult items, is well documented and leads to what are called "difficulty factors", extra factors that appear due to the asymmetry of distributions of items (Embretson & Reise, 2000; Ferrando & Anguiano-Carrasco, 2010; McDonald & Ahlawat, 1974).

If the items are polytomous (Likert type), as we have already anticipated, the approximation to the ideal continuity condition is reasonably adequate when the number of response options is 5 or above, and again the distribution of items is approximately normal. In this case, asymmetry would also be a problem since linear relationships between items measuring the same trait may be attenuated and even brings about non-linear relationships (see West, Finch, and Curran, 1995). Another factor to consider is item discrimination. The items with intermediate discrimination indices are not a problem, but when discrimination indices are extreme, the Pearson correlation coefficient is not adequate.

Nevertheless, these classic recommendations have been systematically ignored in key works focused on updating the decision criteria used in EFA (Conway &Huffcutt, 2003; Fabrigar et al., 1999; Henson & Roberts, 2006; Park, Dailey, & Lemus, 2002; Pérez & Medrano, 2010). These studies do not contemplate the type of items to be analyzed nor the type of matrix to be used. In these works, only the possibility of analyzing Pearson's matrix of product-moment correlations has been considered, which, as we have seen is not always the best choice. This, together with the fact that SPSS, the most commonly used software, in its original configuration does not provide the opportunity of analyzing another type of matrix other than the product-moment correlation matrix (or the variance-covariance matrix) which has led to the usual practice of calculating Pearson's product-moment correlation matrix on ordinal, dichotomous or polytomous items, without previously studying item distribution.

Current recommendation

At present, reviewing item distribution as a previous step in the analysis is still recommended, but other types of association matrices can be used if convenient :polychoric(for polytomous items) or tetrachoric(for dichotomous items) (Bandalos & Finney, 2010; Brown, 2006; Ferrando & Anguiano-Carrasco, 2010). As on other occasions, the approximation to normality demanded by the polytomous variables is greater or lesser depending on the author. Some recommend distributions with skewness and kurtosis coefficients in the range of (-1, 1) (e.g., Ferrando & Anguiano-Carrasco, 2010; Muthén & Kaplan, 1985, 1992). Others however, consider as acceptable values in the range of (-1.5, 1.5) (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009), or even the range (-2, 2) (Bandalos & Finney, 2010; Muthén & Kaplan, 1985, 1992). There is agreement that the negative impact of asymmetry interacts with other factors such as the sample size or the amount of items defining each factor (Forero et al., 2009; West et al., 1995) so approximation to normality demand must grow as other conditions become more unfavorable.

It is recalled here that an EFA based on the polychoric correlation matrix will require larger samples than EFA to reach the same degree of precision and stability based on Pearson's product-moment correlation matrix, therefore if the sample is small (200 subjects) and the distributions are adequate, factor analyzing Pearson's correlation matrix is recommended. When in doubt, we can clarify by performing and comparing the analysis of both matrices.

Specific software allowing analysis of the most adequate correlation matrix are LISREL (Jöreskog & Sörbom, 2007) and MPlus (Muthén & Muthén 1998-2012). There is also specific and free software, such as FACTOR (Lorenzo-Seva & Ferrando, 2006) or the "Psych" package in R (Revelle, 2014). In addition, SPSS users have TETRA-COM (Lorenzo-Seva & Ferrando, 2012a), a program for SPSS that estimates the tetrachoric correlation matrix. A routine developed by Basto and Pereira (2012) has recently appeared also allowing implementing polychoric correlation matrices in SPSS analysis.

In any case it is also advisable to analyze the bivariate distributions of each pair of items, as recently mentioned by Pérez and Medrano (2010) to identify patterns of non-linear relationships between items. These patterns would violate the linear assumption of EFA and again add noise to the matrix analyzed and confusion to the factor structure identified.

Adequacy of data to Factor Analysis

Regardless of the matrix to be factor analyzed, its degree of adequacy to FA must be checked. One common way is through Kaiser Meyer Olkin's KMO Test for Sampling Adequacy (Kaiser, 1970). This reflects the influence of all factors: size of correlations between items, sample size, number of factors and items. This measure of adequacy indicates the size of the correlations between the measured variables. If the correlations are sufficiently large, the matrix is adequate for its factorization as it will provide stable results, replicable in other samples, regardless of sample size, number of factors or items. If KMO is large enough, the results will not be fortuitous. Kaiser considered a matrix with KMO values below .50 inadequate for FA; mediocre if KMO values ranged between .60 and .69; and satisfactory only for values of .80 upwards. However, in our experience many authors have only collected the first of these values (.50), as a cut-off point (see Ferguson & Cox, 1993; Hair, Anderson, Tatham, & Black, 2005; Tabachnick & Fidell, 2001). Other authors prefer to increase the standard to .70 and even .80 (Costello & Osborne, 2005; Ferrando & Anguiano-Carrasco, 2010). The results of the empirical study presented in the second part of this paper support this value.

What is the most appropriate factor estimation method?

The recommended estimation methods are commonly Maximum Likelihood (ML) and Ordinary Least Squares (OLS), though the latter actually agglutinates a set of methods, notably principal axes and unweighted least squares based on the same principle: to minimize the sum of squares of differences between the correlations observed and those reproduced by the model; i.e. to make residuals as close to 0 as possible (Ferrando & Anguiano-Carrasco, 2010).

Maximum Likelihood: This method is inferential (Lawley & Maxwell, 1971). It is a factor estimation method providing the parameter estimates most likely to have produced the observed correlation matrix if the sample proceeds from a normal multivariate distribution with *m*latent -factors. The correlations are weighted by the inverse of the uniqueness, and an iterative algorithm is used for the estimation of parameters. This method has the advantage of allowing testing the goodness-of-fit of the model to the data, through an index that follows a chi-square distribution, and provides the standard error and significance tests around the estimated parameters.

One disadvantage of ML is that it requires compliance with the multivariate normality assumption. Some authors, such as Finney and DiStefano (2006), recommend testing this assumption of multivariate normality - as does the FACTOR program (Lorenzo-Seva & Ferrando, 2006) and the Structural Equation Model (SEM programs). Yuan and Bentler (1998) are less strict, and suggested this assumption is unrealistic in psychology. There is abundant literature based on simulation studies showing that the ML method is robust in violating this assumption when variables have an approximately normal univariate distribution (e.g., Forero et al., 2009; Muthén & Kaplan, 1985, 1992; West et al., 1995). Another drawback is that the chi-square index is very sensitive to sample size (Brown, 2006; Tabachnick & Fidell, 2001). In addition, Ferrando and Anguiano-Carrasco (2010) remind that the chi-square test assumes that the proposed model with *m*factors fits the population perfectly, therefore, that all error is sampling error, ignoring approximation error (the degree to which the model is a reasonable approximation to what occurs in the population, given the data seen in the sample). All this leads to the rejection of models that do represent a good approximation to the latent factor structure, in favor of models with more factors than those that have theoretical meaning, i.e. overfactorized models. Thus, in practice, other goodness-of-fit indicators derived from the chi-square test evaluating the approximation error and the degree of fit of the model, are usually considered. These will be seen in another section.

Therefore

A) The ML method and factor analyzing Pearson's Product-Moment Correlation matrix are recommended, if the items have sufficient response categories (5 or more), or are continuous (unlikely), and reasonably comply with the normality assumption , so that the linear relation assumed in bivariate relations between items can be observed (Flora, LaBrish & Chalmers, 2012). It is also recommended considering other goodness-of-fit indices derived from the chi-square test that evaluate the approximation error and the degree of fit of the model to the data.

B) Analyzing the polychoric correlation matrices by ML is not recommended. Although the parameter estimates are generally unbiased (Bollen, 1989), and even reproduce the measurement model better than the analysis of product-moment correlations when some items do not meet the assumption of normality (Holgado-Tello, Chacón-Moscoso, Barbero, & Vila-Abad, 2010) the goodness-of-fit tests based on chi-square, as well as the standard errors (and therefore the significance tests of the parameters involved) will be biased (Bollen, 1989; Satorra & Bentler, 1994), and, consequently, interpretation of these should be avoided. When analyzing the polychoric correlation matrix, it is recommended using OLS (e.g. Flora et al., 2012, Forero et al., 2009; Lee, Zhang, & Edwards, 2012).

Ordinary Least Squares: OLS methods gather a series of descriptive methods with the common denominator of determining the factor solution making residuals as close to zero as possible. These methods have shown good results in the factorization of ordinal items when analyzing the polychoric correlation matrix (Forero et al., 2009; Lee et al., 2012).

Among these methods, Principal Axis has been the classic recommended option when the normality assumption is not met, which is more likely as the number of response categories is reduced (Fabrigar et al., 1999). This method can be applied in a non-iterative way, and the most common estimates of commonalities are the squared multiple correlations between each observed variable and the rest. These substitute the diagonal values of the correlation matrix, producing what is called the reduced correlation matrix, the input for the factor analysis. However, in most recent versions of programs, the principal axis method is applied iteratively by default. On estimating the initial commonalities, the reduced matrix is decomposed into its eigenvalues and eigenvectors; the latter are rescaled and form the first iteration factor matrix .From this initial factor matrix, the commonalities are reestimated, replacing the initial ones.

This iterative process continues until all differences between the commonalities of two successive iterations are so small that the convergence criterion or a maximum number of iterations are reached. The aim is to obtain the best possible commonality estimates based on the number of factors retained. The adequacy of the principal axis method depends on the quality of the estimates of communalities. When applied iteratively, principal axis shows more similar results to those provided by other methods such as the Unweighted Least Squares (ULS) method (Joreskög, 1977) (compared to the non-iterative approach). Rather than using the reduced matrix as input, with the communalities estimated on the diagonal, ULS minimizes the sum of the squares of the differences between the observed and reproduced correlation matrices. This method is the most popular today (see Flora et al., 2012), as it performs well with small samples even when the number of variables is high, especially if the number of factors to retain is small (Jung, 2013). It also prevents the occurrence of Heywood cases (factor loadings greater than unity and negative error variances), which are more common with other estimation methods. It should be noted that ULS is virtually interchangeable with the Minimum Residual (Minres) Method (Harman & Jones, 1966; Joreskög, 1977). Depending on the software used, we will have either estimation method available.

Ferrando and Anguiano-Carrasco (2010) stated that "in a situation where (a) variables have acceptable distributions, (b) the solution is well determined, and (c) the proposed model is reasonably correct, both OLS and ML solutions will be almost identical. In this case, ML has the advantage of offering additional useful indicators in the assessment of fit." (p. 28) Otherwise, there will be convergence problems; unacceptable estimates and unreliable indicators. In these cases the authors recommend using OLS. It should be added that though some robust estimation methods have been suggested for violation of the multivariate normality assumption (e.g., Weighted Least Squares (WLS), Robust Weighted Least Squares (WLSMV), or Robust Maximum Likelihood), these have been applied more in confirmatory than exploratory models. Programs such as Mplus do have some of these methods implemented for both CFA and EFA. These robust methods (the choice of a specific method depends on sample size) are the most recommended in structural equation models in general, and in CFA, in particular, when analyzing ordinal data that depart from the norm (e.g., Curran, West, & Finch, 1996; Flora & Curran, 2004). However, with EFA, we must still perform simulation studies that allow comparing the advantages of these robust methods to others such as ULS (beyond the possibility of offering goodness of fit indices) or even ML, when the data follow a distribution that departs from the norm. There is also a case for Bayesian estimates (Muthén & Asparouhov, 2012).

How to select the proper number of factors

This is perhaps the most important aspect of an EFA. The number of common factors required to explain the relationships between items, and the composition of these factors, are the two central issues in the interpretability of the factor structure obtained in the analysis. If fewer factors than needed are retained, the resulting factor loading patterns are harder to interpret, and the identified factors are confusing. If more factors than needed are retained, latent variables are "fabricated" with little theoretical or substantive meaning. The decision is important, and perhaps why it is one of the aspects where more options have been provided to make the choice. We shall see.

Classic recommendation

The classic recommendation is to use the Kaiser rule: we will select factors with eigenvalues greater than 1, extracted from the original correlation matrix (and not the reduced matrix, i.e. the matrix with the commonalities in the diagonal). The Kaiser rule has been routinely used in the package known as "Little Jiffy" which includes the application of PCA with the Kaiser rule and the Varimax rotation method. This was suggested by Kaiser in 1957 as a computationally efficient alternative to the most appropriate but computationally less efficient options based on EFA and oblique rotations. There have been many changes since 1957 for us to continue using the same procedures. Nevertheless, studies reviewing the way EFA has been applied in research from 1975 to now, showed how this criterion is the most used, either alone or in combination with another (Conway & Huffcutt, 2003; Fabrigar et al. 1999; Henson & Roberts, 2006; Park et al., 2002; Pérez & Medrano, 2010) both when using PCA and EFA. Fortunately, recent studies show that the frequency with which this rule is used as the sole criterion is wavering toward an increase in cases used in combination with other criteria (Conway & Huttcuff, 2003; Henson & Roberts, 2006).

The most frequently combined criterion with the Kaiser rule is the Scree test (Cattell, 1966). This is used with EFA, but actually what is analyzed as in the case of the Kaiser rule is normally the original correlation matrix, not the reduced one (Ford et al., 1986).

The other classical criteria less often used in combination with the Kaiser rule are: 1) the solution offering the best possible interpretation, and 2) the number of factors based a priori on theory (Conway &Huffcutt, 2003; Fabrigar et al., 1999; Ford et al., 1986).

Current recommendation

Although Kaiser's rule is the most inadvisable of all possible options, it is still the most widely used (Costello & Osborne, 2005). For example, Lorenzo-Seva, Timmerman, and Kiers (2011) excluded this criterion from their simulation study as it was clearly inappropriate. In this regard, Ferrando and Anguiano-Carrasco (2010) stated that one of the main drawbacks of this procedure is that the number of factors it identifies is directly related to the number of items analyzed; if *n* variables are analyzed, the number of factors that this rule obtains will oscillate between n/5 and n/3, regardless if the scale is unidimensional. Similarly, Lorenzo-Seva et al. (2011) excluded the Scree test, because the subjectivity for its application is difficult to program, so it cannot be introduced in a simulation study in the same way they performed. However, Fabrigar et al. (1999) supported its use when the underlying common factors are clearly defined.

It is strongly recommended that the decision on the number of factors be made with the following in mind: 1) several objective criteria, and always considering 2) the interpretability of the solution obtained, and 3) the theoretical background (Lorenzo-Seva et al., 2011).

The available objective criteria vary greatly depending on which software is used, and to a lesser degree on the factor estimation method.

Parallel Analysis (PA) selects common components or factors that present eigenvalues greater than those obtained randomly (Horn, 1965). If the analyzed matrix came from a population where items were not related, what eigenvalues could reach the common factors extracted from that matrix? And how many common factors obtained on the actual analyzed matrix exceed these "spurious" eigenvalues? The answer provides the correct number of common factors. This technique was developed to be used on the original correlation matrix (such as the Kaiser rule or the Scree test) and not on the reduced correlation matrix. However, its use has also been recommended in identifying the number of common factors (Hayton, Allen, & Scarpello, 2004; Lorenzo-Seva et al., 2011; Peres-Neto, Jackson, & Sommers, 2005; Velicer, Eaton, & Fava, 2000).

Despite being recommended in numerous studies, this criterion is rarely used. This is explained by the accessibility of the procedure, since it is not implemented in the SPSS directory (the most commonly used program by researchers in social sciences). Fortunately, there is already a routine in R that allows applying PA and other criteria to identify the number of factors in the context of the SPSS program (Basto & Pereira, 2012), programs for SPSS and SAS that determine the number of components through PA and Vellicer´s MAP (O'Connor, 2000), and a program specifically designed for PA on the Internet: ViSta-PARAN (Ledesma & Valero-Mora, 2007). FACTOR (Lorenzo-Seva & Ferrando, 2006) is freely distributed and also implements the latest advances in this and other aspects.

Three currently recommended criteria start from a different logic to PA. All are based on the evaluation of residual correlations. If the appropriate number of common factors has been extracted, ideally no common variance should remain and so the residual correlations will tend to zero. These criteria are: 1) direct inspection of the distribution of the standardized residuals (median and extreme values, which will indicate where among variables the problem lies); 2) the root mean square residual (RMSR) and 3) the gamma index or GFI (Tanaka & Huba, 1989). These latter two emerged in the context of a more general framework: SEM.

The RMSR is a descriptive statistic that condenses the information contained in the residual correlation matrix. It is a measure of the average magnitude of the residual correlations. Commonly, Harman's (1980) proposal of a cut off point of .05 has been used to consider an acceptable fit. Lorenzo-Seva and Ferrando (2006) instead recommended the criterion initially proposed by Kelley (1935); Kelley used the standard error of a correlation of zero as the reference value in the population from which the data originated, approximately $1/\sqrt{N}$ (see Ferrando & Anguiano-Carrasco, 2010). The RMSR can be used with any factor estimation method, though it is not implemented in generalist software such as SPSS.

The GFI is a normed measure of goodness of fit ranging from 0 to 1 and can be used with most factor estimation methods, although it is uncommon in generalist programs such as SPSS. It indicates the proportion of covariation between items explained by the proposed model, and is therefore a species of multivariate coefficient of determination. Values above .95 are indicators of good fit (McDonald, 1999).

The RMSEA (Steiger & Lind, 1980) is an index based on the chi-square statistic, so can only be obtained with factor estimation procedures which offered it, which depends on the software used (e.g. FACTOR includes it for ULS and ML, but is not available in generalist programs like SPSS, although it can be calculated by the user). We find recommendations for it in works by Browne and Cudeck (1993), Fabrigar et al. (1999), Ferrando and Anguiano-Carrasco (2010), and Lorenzo-Seva et al. (2011). This index estimates the approximation error of the proposed model. It is an index relative to the degrees of freedom of the model. Thus, it can favor the selection of more complex models. Values below .05 are considered excellent, while those above .08 indicate insufficient fit.

Finally, another criterion applied in the context of the PCA which has been generalized to the EFA, is the Minimum Average Partial (MAP) Test (Velicer, 1976). This procedure focuses on identifying the amount of components providing the minimum partial correlation between the resulting residuals. Zwick and Velicer (1986) indicated that this procedure was adequate in identifying the number of components. It has also been applied in the identification of the number of common factors. This criterion is applied in the R command that allows improving the SPSS features for EFA (Basto & Pereira, 2012), in the program developed by O'Connor (2000), and also in the program FACTOR (Lorenzo-Seva & Ferrando, 2006).

As for the interpretability of the solution, it is essential evaluating what the objective indicators suggest. It is little use that a model with two factors fits better than another with 3, if the third factor is poorly defined (with fewer than 3 items with factor loadings greater than .30, for example), or cannot be interpreted due to a lack of meaningful content. We must highlight the trend over the last decades (Fabrigar et al., 1999, Lorenzo-Seva et al., 2011) to distinguish between common major and minor factors. The former must be retained as they explain a substantive part of the items comprising the scale. The minor common factors also explain a part of the common variance, but only a small part, which does not become interpretable in the context of what we wish to measure with that set of items. The example of Lorenzo-Seva et al. (2011) is clear: if two items measuring different personality traits have as a common denominator where the person is situated. Perhaps this common part produces a certain covariance among individuals´ responses, and that a minor common factor emerges to explain this covariance .That common factor will have circumstantial meaning but not substantive meaning. Consequently, the current recommendation is not to explain the greatest amount of common variance as possible, but most of the common variance possible to explain with the proper number of common factors, which will be those factors that have meaning. This is as some readers may have surmised the eternal dilemma between parsimony and plausibility.

One criterion not recommended is interpretation of explained variance, as it is not a satisfactory indicator of the adequacy of the number of identified common factors. The common item variance cannot be distributed among the various common factors (except under an estimation method: Minimum Rank Factor Analysis). If so, some eigenvalues would be negative, as otherwise the sum of the common variance explained by each factor may exceed the total common variance, precisely because it is common, and has been accounted for in the common part of the variance of different items. A more detailed explanation can be found in Lorenzo-Seva (2013). Bandalos and Finney (2010) can also be consulted.

In summary, it is recommended using a mix of objective criteria and criteria based on the theory and interpretability of the solution, and use this combination to compare alternative solutions. It would be ideal to present the most plausible and parsimonious combination. It should be remembered that the program used for analysis, along with the factor estimation method, will greatly limit the specific type of objective criteria used. It is advisable to abandon those programs that do not allow use of the best methods, unless a macro is known that solves this problem. The Kaiser rule which is the most used criteria in all the reviewed studies is discouraged; and we recommend incorporating procedures based on: 1) the PA, 2) the MAP method or 3) the RMSEA goodness-of-fit index, if possible. Beyond these, no further universal recommendations can be made.

The interested reader can find more detailed information about these and other methods and their functioning in a simulation study carried out by Lorenzo-Seva et al. (2011).

What kind of factor rotation and what criteria for item assignment must be employed?

Following the factor estimation phase, the solution is rotated to achieve the greatest simplicity and interpretability.

Classic recommendation

Thurstone (1947) suggested that factors be rotated in a multidimensional space to obtain the solution with the best simple structure. The factor rotation can be orthogonal or oblique. The former method assumes independence of factors; while the latter allows correlation between factors.

The use of either type of rotation has practical implications when presenting the EFA results. The contribution of a particular item to a given factor is indicated both with factor pattern coefficients and factor structure coefficients. The former are found in the pattern matrix and the latter in the structure matrix. In EFA, the structure matrix provides correlations between observed variables (items) and latent variables (factors). When orthogonal rotation is used, since it is assumed that the factors are uncorrelated; the structure matrix and the pattern matrix are exactly the same, therefore it is enough to provide the pattern coefficient matrix. However, when oblique rotation is used, the factor correlation matrix is not an identity one, so the structure and the configuration matrices will not offer the same coefficients. In oblique rotation, both matrices must be provided, and the interpretation of results must be done by first considering the structure matrix coefficients, and then examining the pattern matrix coefficients (Courville &Thompson, 2001; Henson & Roberts, 2006; Gorsuch, 1983; Thompson & Borrello, 1985).

The most widely known rotation criteria available in most commercial statistical analysis programs are: Varimax (Kaiser, 1958) for orthogonal rotation, and direct Oblimin (Clarkson & Jennrich, 1988) and Promax (Hendrickson &White, 1964) for oblique rotation. Varimax has been suggested as an orthogonal rotation criterion when there is no dominant factor, whereas Quartimax has been proposed as an alternative orthogonal criterion when a single general factor is expected (Carroll, 1953). Equamax combines both criteria (single factor / several factors) and offers intermediate solutions. Weighted versions of the Varimax criterion have also been formulated (e.g., Cureton & Mulaik, 1975)

In oblique rotation, the most popular criterion has been direct Oblimin. The Quartimin criterion (Jennrich & Sampson, 1966) is equivalent to the Quartimax criterion, but in oblique rotation. Other criteria have recently been proposed, some quite new to the usual EFA users, such as Geomin (Yates, 1987), Promin (Lorenzo-Seva, 1999) and weighted Oblimin (Lorenzo-Seva, 2000). For a detailed presentation of different rotation criteria we recommend Browne (2001), and Sass and Schmitt (2010).

Although different rotation criteria have traditionally been formulated under a certain method (orthogonal or oblique), more recent software developments show that the same criterion may be available using both orthogonal and oblique methods (e.g., Mplus, Muthén & Muthén, 1998-2012).

Historically, researchers have selected the rotation criterion for use in EFA depending on the popularity of a specific criterion at the time (e.g., Varimax, an integral part of the "Little Jiffy pack"), or depending on the recommendation for using orthogonal rotation when factors are independent, and oblique rotation when factors are related.

Current recommendation

Over the last twenty years, studies reviewing the use of EFA have shown an evolution from major use of orthogonal rotation (specifically the Varimax criterion), to more common use of oblique rotation. Ford et al. (1986) concluded that, in the reviewed studies, orthogonal rotation was mostly applied. Specifically, Ford et al. (1986) found that about 80% of the reviewed EFAs used orthogonal rotation, while 12% either used oblique rotation or did not rotate the factor solution (the remaining 8% provided no information). Included in this study are those who use oblique and who do not rotate the solution; in the same category; thus we conclude that the percentage of analysis using oblique rotation was less than 12%.

Fabrigar et al. (1999) came to a similar conclusion as Ford et al. (1986), but with an increasing trend toward oblique rotation. Results indicated that 48.3% of EFAs evaluated used orthogonal rotation (particularly Varimax), and 20.6% oblique rotation.

Conway and Huffcutt (2003) recently reconfirmed the decreasing trend of orthogonal rotation. The percentage of EFAs analyzed using this rotation had dropped to 41.2%. However, the increase in the percentage of use of oblique rotation could not be confirmed. Their study found it was used in only 18% of cases. Contrary to expectation, this result indicated a slight decrease in oblique rotation. Conway and Huffcutt (2003) explained this by referring to the percentage of works not indicating the rotation used. In fact, in their study, this percentage was 18%, while in the study by Fabrigar et al. (1999) it was only 8%.

This evolution in the use of rotation criteria is because orthogonal rotation is traditionally thought to produce simpler and more easily interpretable structures (e.g., Nunnally, 1978). However, different studies (Fabrigar et al., 1999; Finch, 2006; Henson & Roberts, 2006; Matsunaga, 2010; Park et al., 2002; Preacher & MacCallum 2003) show this not to be so, but that indeed oblique rotation can present clearer, simpler and more interpretable structures.

These studies suggested using oblique rotation regardless of the underlying theoretical model (independent or related factors). There are several arguments for this: 1) almost all phenomena studied in the social and health sciences are roughly interrelated, so it is difficult to find perfect orthogonal solutions. It follows that imposing an orthogonal factor solution could be unrealistic; 2) If the construct under study really presents a structure of independent factors, this orthogonality will be reflected in the results (by allowing correlations between factors using an oblique approximation, the obtained correlations will be low); 3) Finally, if correlations between factors were consistently low (below .30 or .20), it is suggested repeating the analysis using an orthogonal solution. If both solutions were similar, considering the criterion of parsimony; it would be advisable to provisionally accept the orthogonal solution (Ferrando & Anguiano-Carrasco, 2010).

The recommendation is to start by applying oblique rotation criteria. But from those formulated, which is best? Perhaps this question is unanswerable. It has recently been reported that the number of possible factor structures hypothesized is very broad and so the rotation criteria applied in each particular case will be different (e.g., Browne, 2001). In this line, the work of Sass and Schmitt (2010) tried to compare different rotation criteria in different situations (e.g., perfect simple structure, approximate simple structure, complex structure, general structure) and provide guidelines for performing an EFA. To that end, these authors performed a simulation study with Mplus where they evaluated different oblique rotation criteria (Quartimin, CF-Equamax, CF-Facparsim and Geomin) and their ability to reproduce different factor pattern matrices. Sass and Schmitt (2010) concluded there is no definitive answer to what rotation criterion produces the "best" solution. Apparently, there are no correct or incorrect rotation criteria, nor any that produce better or worse solutions. In contrast, one must be aware that selecting ne rotation criterion or another can have important effects on estimated factor pattern loadings and between-factor correlations. The different rotation criteria, under certain circumstances, can lead to very similar factor pattern matrices; but under other circumstances may produce different and even contradictory ones. It is in these situations that the researcher must make the difficult decision to choose the most appropriate rotation criterion, i.e., that providing the simplest and most informative solution (Asparouhov & Muthén, 2009). Ultimately, 'the selection of "best" rotation criterion must be made by the researcher" (Sass & Schmitt, 2010, p. 99). This suggestion was already made by Browne (2001, p. 145): *"the choice of the best solution therefore cannot be made automatically and without human judgment"*. The researcher can test different factor solutions using different rotation criteria, and based on results, select the "best" rotation criterion offering the simplest and most informative factor solution. It is even advised, that as well as providing a justification of the chosen rotation criterion, the different "factor pattern matrices" from different rotation criteria" are offered, so "this allows the reader to draw their own conclusions based on the competing factor structures" (Sass & Schmitt, 2010, p. 97).

Regarding the criterion of assigning items to factors, another aspect that can hugely vary the interpretation of the solution obtained, common practice is to retain factor loadings above .30 or .40 (Bandalos & Finney, 2010; Guadagnoli & Velicer, 1988; McDonald, 1985). In fact, Tabachnick and Fidell (2001) stated that .32 could be a good general rule in the minimum loading to be considered, equivalent to approximately 10% of the explained variance. Other authors are slightly stricter and place the cutoff point at a minimum of .40 (MacCallum et al., 1999; Velicer & Fava, 1998; Williams, Brown, & Onsman, 2010) and also recommend raising this cutoff point if the sample is lower than 300 cases. It is also recommended that the discrepancy between factor loadings in the first two factors be .50/.20 or .60/.20 (i.e. a difference of .30-.40). Some studies, particularly empirical ones, use the .60/.40 discrepancy (Henson & Roberts, 2006; Park et al., 2002).

Items that do not surpass the criterion or set of established criteria must be revised in both their aspects: substantive and methodological, to identify why they do not work well. We can then assess whether to remove them from the test, modify them somehow for inclusion in a new version of the test, or if new items of similar content must be added to adequately sample the content of the factor to be measured (Costello & Osborne, 2005), leading us to re-examine the aforementioned content validity (Bandalos & Finney, 2010). In addition, after eliminating these items, a new factor analysis will be carried out with the reduced scale. Ideally the analysis would be repeated after eliminating one of the inappropriate items one by one (Bandalos & Finney, 2010). Small variations such as eliminating a pair of items can often substantially modify the final result of the analysis.

As the reader may have already noticed, the aspects discussed so far are many and varied, and yet do not touch on all that should be dealt with in the context of EFA. We have only tried to present those most relevant that the applied researcher must solve when beginning work. We have omitted, among others, non-linear FA, on which there is extensive literature, the issue of factor scores, which can be consulted in the chapter on Exploratory Factor Analysis of Abad et al. (2011) and neither have we have provided recommendations on the use of item parcels, which can be consulted in West et al. (1995) and in a recent and extensive work by Little, Rhemtulla, Gibson and Schoemann (2013). And finally, we have not provided recommendations regarding when and how to use the originally entitled "Extension Analysis" (Gorsuch, 1997 a, b), for which the EXTENSION program exists, compatible with SPSS, SAS And MATLAB (O'Connor, 2001).

We refer interested readers to the sources cited for information on these aspects. We will now summarize all the above recommendations in a manner useful to the reader.

Exploratory Factor Analysis of Items: Brief User Guide

We have presented a lot of information thus far, and at this point the aim of offering clear and current recommendations may have blurred. Thus, in conclusion, the most important current recommendations will be summarized and integrated into a brief guide around four central aspects: 1) the type of data and the association matrix; 2) the factor estimation method; 3) the number of factors and finally, 4) rotation method and item allocation.

1) Type of data and association matrix

Recommendation

The applied researcher intending to perform an EFA to evaluate the scale of interest should use appropriate sample sizes, be consistent with the ordinal, polytomous (Likert type) or dichotomous nature of the items to be analyzed, use appropriate software in each case, and check the adequacy of their data to the FA at least the KMO test.

As a general rule, Pearson's correlation matrix is limited to the case where items are continuous, or if not, have five or more response alternatives and approximately normal distributions (a demanding criterion uses values for kurtosis and skewness in the range of -1 to 1). In most other cases it is recommended using the polychoric or tetrachoric correlation matrix, depending on the case.

The exception occurs if the sample is small (200 subjects or fewer). In this case, the polychoric correlations might not be very stable and so less recommendable than the Pearson correlations, therefore we suggest relaxing this requirement and using values for greater asymmetry and kurtosis (-2, 2). When in doubt, the researcher can compare the solutions obtained with both matrices and decide on the best solution.

Sample size is not easy to anticipate, as it depends on the psychometric characteristics of items, of the type of association matrix, and these two aspects are related. However, if we wish to have some guarantee of being able to use the polychoric / tetrachoric correlation matrices we should not use samples with fewer than 300 subjects. With the Pearson correlation matrix, a minimum sample of 200 cases is recommended, only valid as a starting point. Where factor loadings are low (less than .40) and the number of items per factor is also low (3 items) larger sample sizes are required as a guarantee of generalization of results. As it is customary to use convenience samples, two problems must be considered: non-representativeness and the attenuation and restriction of variance.

The adequacy of the data to the FA is considered "sufficient" when the KMO measurement takes values between .70-.79, and "satisfactory" with values greater than .80.

2) Factor estimation method

Recommendation

The informed researcher must adapt the method to the type of data to be analyzed: if the items are ordinal but have approximately normal distributions, - in this case the criterion is to accept absolute values below 2 for asymmetry and kurtosis, then the appropriate method is ML applied to the Pearson correlation matrix as this provides more information (goodness-of-fit indices, significance tests and standard errors around the estimated parameters).

Otherwise, it is recommended applying an OLS-based method such as ULS. This allows analyzing matrices in adverse situations, even with few cases and many items, and without the need to make distributional assumptions. This method is also recommended when the ML solution is inadequate as it shows convergence problems or Heywood cases, or other similar anomalies. In these cases, we must carefully assess if these estimation problems are not masking data that do not actually fulfill required assumptions (such as normality or linearity) or a poorly specified model (Fabrigar et al., 1999).

PCA is the estimation method which is strongly discouraged, a method to explain the variance of each individual item. It is not a suitable method for reaching the aim of FA, which is to explain correlations among items from the identification of a set of common factors. It should not even be used with an oblique rotation of identified components.

3) Number of factors to retain.

Recommendation

As regards the number of factors to retain, the Kaiser's criterion is also an option as strongly discouraged as the PCA estimation method. An informed researcher must employ a range of methods to decide: they must retain only "major" common factors considering the number of items (minimum 3-4), the factor loading size (minimum .40), and the meaning of items that define them, and must apply at least one "extra" objective criterion in the margin of those offered by a conventional EFA - which are Scree test, GFI or RSMR -, or an EFA by means of SEM- we have mentioned RMSEA as a criterion, although there are others which can be used. When the proportion of items per factor is small, and the sample size is 200 cases or more, this "extra" objective criterion can be a Parallel Analysis or the MAP method. Bearing this information in mind the recommendation is as follows:

1) to obtain results about the goodness-of-fit of the model with the number of factors fixed to the expected number,

2) to obtain results about the goodness-of-fit of other models that may appear plausible following initial analysis, and

3) to compare and decide.

The use of the percentage of variance explained is strongly discouraged as a criterion, as it is confusing.

4) Rotation method and item assignments

Recommendation

Oblique rotation is highly recommended, even when it may seem appropriate, as they obtained result will highlight. However, the same cannot be said for orthogonal rotations. What is unclear is the most appropriate oblique rotation method, since there are no clear criteria. The researcher must try several and choose which presents better interpretability.

As for the criterion for interpreting the factor loading of an item, the recommendation is also clear: never below .40, especially if the sample is fewer than 300 cases.

Items not exceeding this value should be eliminated from the analysis and subjected to a substantive and methodological examination to choose from three options: eliminate or revise the item or add new ones sampling the facet related to that particular item. In any case, we should reexamine the adequacy of the test content to the construct to be measured. In addition, a new factor analysis with the reduced scale will be performed after eliminating one of these items each time.

Let us stress that use of the "Little Jiffy" pack is discouraged: PCA plus Kaiser plus Varimax. For some time it has been the most effective way of summarizing the observed relationships between test items. Even if under certain unlikely assumptions it could provide adequate results; this is not the case today.

Conclusion

In an article by Martinez and Martinez (2009, p. 373) on the new conceptualizations of validity - and the difficulty with which they are taken up by the psychometric community -, conclude with the following example: "Jacob Cohen, one of the most influential researchers in methodology, has spent more than forty years defending the use of effect size indices rather than the classic "p-value," with incredibly logical arguments" Cohen (1990) explains, in spite of everything. "The problem is that, in practice, current research hardly pays attention to the effect size" (Cohen, 1990, p. 1310). Sedlmeier and Gigernzer (1989) concur with this in their article "Do studies on statistical power have an effect on the power of studies?", that the power of published studies is, on average, insufficient .44. The use of statistical power may have increased in recent years, but it is not common practice to report it. If something as simple as the study of the effect size when interpreting analyses is still not routinely applied, despite the large number of studies showing its suitability, it is not surprising at what occurs in the more complex context of EFA.

Borsboom, in his article "The Attack of the Psycometricians" (2006), also concludes that psychometric advances seem not to have contributed much to the advance of psychology in the last 50 years. And it is not for lack of advances in psychometrics but for lack of application of those advances in psychology. We believe that the key is, as Cohen (1990, p. 1309) puts it, "that the process by which one conceives, plans, executes, and writes research must depend on *the informed judgment of one's self as a scientist*" (our italics).

It is not a question of repeating what other researchers have done before; we cannot advance in this way. It is about being informed and having the judgment to take the best decisions. This is how we leave the maze, through the front door as we said at the beginning of this paper. And that should be the maxim of all those who work in the world of research: researchers, editors, reviewers and professors. This is our small contribution.

References

Abad, F. J., Olea, J., Ponsoda, J., & García, C. (2011). *Medición en ciencias sociales y de la salud [Measurement in social sciences and health]*. Madrid. Síntesis.

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

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411-423.

Arrindell, W. A., & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, *9*, 165-178.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397-438.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock and R. O. Mueller (Eds.), *Reviewer's guide to quantitative methods*. Routledge: New York.

Barrett, P. T., & Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality Study in Group Behavior, 1*, 23-33.

Basto, M., & Pereira, J. M. (2012). An SPSS R-Menu for ordinal factor analysis. *Journal of Statistical Software, 46*, 1-29. [Consulted in February, 9, 2014]. Retrieved from: http://www.jstatsoft.org/v46/i04/paper.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, *18*(6). Retrieved from: http://pareonline.net/getvn.asp?v=18&n=6.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*, 305-314.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*, 425-440.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203-219.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford Press.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36,* 111-150.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen y J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Carroll, J. B. (1953). An analytic solution for approximating simple structure in factor analysis. *Psychometrika, 18*, 23-28.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Clarkson, D. B., & Jennrich, R. I. (1988). Quartic rotation criteria algorithms. *Psychometrika*, *53*, 251-259.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*(12), 1304-1312.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis.* Hillsdale, NJ: Erlbaum.

Conway, J. M., & Huffcutt, A. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods, 6*(2), 147-168.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis . *Practical Assessment Research & Evaluation, 10*(7), 1-9. [Consulted in January, 21, 2014]. Retrieved from: http://pareonline.net/getvn.asp?v=10&n=7.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: Beta is not enough. *Educational and Psychological Measurement*, *61*, 229-248.

Cureton, E. E., & Mulaik, S. A. (1975). The weighted varimax rotation and the Promax rotation. *Psychometrika*, *40*, 183-195.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29.

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, *14*, 370-388.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

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.

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment*, *1*(2), 84-94.

Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en Psicología [Factor analysis as a technique of research in Psichology]. *Papeles del Psicológo*, *31*(1), 18-33.

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions. *Psicologica, 21*, 301-323.

Ferrando, P. J., & Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory.* Technical Report. Department of Psychology, Universitat Rovira i Virgili, Tarragona.

Ferrando, P. J., Varea, M. D., & Lorenzo, U. (1999). Evaluación psicométrica del cuestionario de ansiedad y rendimiento (CAR) en una muestra de escolares [A psychometric study of the Anxiety of Performance Questionnaire (CAR) in primary school students]. *Psicothema, 11*(1), 225-236.

Finch, H. (2006). Comparison of the performance of varimax and promax rotations: Factor structure recovery for dichotomous items. *Journal of Educational Measurement*, *43*(1), 39-52.

Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation models. In G. R. Hancock and R. O. Mueller (Eds.), *A second course in structural equation modeling* (pp. 269 - 314). Greenwich, CT: Information Age.

Flora, D. V., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9,* 466-491.

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.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, *39*, 291-314.

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

Gorsuch, R. L. (1983). *Factor analysis* (2nd. ed.). Hillsdale, NJ: Erlbaum.

Gorsuch, R. L. (1997a). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68*, 532-560.

Gorsuch, R. L. (1997b). New procedure for extension analysis in exploratory factor analysis. *Educational and Psychological Measurement, 57*, 725-740.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265-275.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Multivariate data analysis*. Prentice Hall International.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago, IL: University of Chicago Press.

Harman, H. H. (1980). *Análisis factorial moderno* [Modern factor analysis]. Madrid: Saltés.

Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (Minres). *Psychometrika*, *31*, 351-369.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*, 191-205.

Hendrickson, A. E., & White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*, 65-70.

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.

Hinkin, T. R. (1995). A Review of scale development practices in the study of organizations. *Journal of Management*, *21*(5), 967-988.

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Munford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, *65*, 202-226.

Holgado-Tello, F. C., Chacon-Moscoso, S., Barbero 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.

Horn, J. L. (1965). A rationale and test for the number of factors in a factor analysis. *Psychometrika, 30*, 179-185.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55.

Izquierdo, I., Olea, J., & Abad, F. J. (2013, September). *¿Se aplica correctamente el análisis factorial exploratorio en la investigación psicológica?* Paper presented to XIII congress of Methodology of the Social Sciences and Health. Tenerife. Spain. Retrieved from: http://13congresometodologia.es/programa/programa-cientifico/programa-cientifico-general/

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, *31*, 313-323.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer-Verlag.

Jöreskog, K. G. (1977). Factor analysis by least-squares and maximum-likelihood methods. In K. Enslein, A. Ralston, and H. S. Wilf (Eds.), *Statistical methods for digital computers* (Vol. 3). New York: Wiley.

Jöreskog, K. G., & Sörbom, D. (2007). *LISREL 8.80*. [Computer Software]. Lincolnwood, IL: Scientific Software International, Inc.

Jung, S. (2013). Exploratory factor analysis with small sample sizes: A comparison of three approaches. *Behavioural Processes*, *97*, 90-95.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187-200.

Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika*, *35*, 401-415.

Kelley, T. L. (1935). *Essential traits of mental life*. Harvard Studies in Education, (Vol. 26). Cambridge: Harvard University Press.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis*. Beverly Hills, CA: Sage.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworths.

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment Research & Evaluation*, *12*(2). Retrieved from: http://pareonline.net/getvn.asp?v=12&n=2

Lee, C.-T., Zhang, G. y Edwards, M. C. (2012). Ordinary least squares estimation of parameters in exploratory factor analysis with ordinal data. *Multivariate Behavioral Research*, *47*, 314-339.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285-300.

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research, 34*, 347-365.

Lorenzo-Seva, U. (2000). The weighted oblimin rotation. *Psychometrika*, *65*, 301-318.

Lorenzo-Seva, U. (2013). *How to report the percentage of explainned common variance in exploratory factor analysis*. Technical Report: Department of Psychology. Universitat Rovira i Virgil, Tarragona.

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: a computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments & Computers*, *38*(1), 88-91.

Lorenzo-Seva, U., & Ferrando, P. J. (2012a). TETRA-COM: a comprehensive SPSS program for estimating the tetrachoric correlation. *Behavioral Research Methods*, *44*(4), 1191-1196.

Lorenzo-Seva, U., & Ferrando, P. (2012b). *Manual of the program FACTOR v. 8.10*. Retrieved from: http://psico.fcep.urv.es/utilitats/factor/Manual-of-the-Factor-Program.pdf.

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*(6), 497-498. [Fecha de consulta: 21 Enero 2014]. Disponible en: http://apm.sagepub.com/content/early/2013/05/07/0146621613487794.citation.

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.

MacCallum, R. C., Widaman, K. F., Preacher, K., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*, 611-637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of Golden Rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's findings. *Structural Equation Modeling, 11*(3), 320-341.

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Measurement fruitfulness of exploratory structural equation modelling (ESEM) in addressing key substantive issues in motivation and engagement research. *Journal of Psychoeducational Assessment*, *29*, 322-346.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471-491.

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85-110.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Applica-

tion to students' evaluations of university teaching. *Structural Equation Modeling*, *16*, 439-476.
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.
Martínez, J. A., & Martínez, L. (2009). El análisis factorial confirmatorio y la validez de escalas en modelos causales [Confirmatory factor analysis and the validity of the measurement scales within a causal modelling framework]. *Anales de Psicología*, *25*(2), 368-374.
Matsunaga, M., (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, *3*(1), 97-110.
McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: LEA.
McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: LEA.
McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement*, *24*, 99-114.
McDonald, R. P. (2005). Semiconfirmatory factor analysis: The example of anxiety and depression. *Structural Equation Modeling*, *12*, 163-172.
McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, *27*, 82-99.
Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling. In Hancock, G. R. and Mueller, R. O. (Eds.), *Structural equation modeling: A second course* (2nd ed.). Charlotte, NC: Information Age Publishing.
Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw Hill.
Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*(2), 159-168.
Muthén, B., & Asparouhov, T. (2010). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313-335.
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.
Muthén, B., & Kaplan D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19-30.
Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.
Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide.* Seventh Edition. Los Angeles, CA: Muthén & Muthén.
Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, *32*, 396-402.
O'Connor, B. P. (2001). EXTENSION: SAS, SPSS, and MATLAB programs for extension analysis. *Applied Psychological Measurement, 25*, 88.
Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, *28*(4), 562-577.
Peres-Neto, P. R., Jackson, D. A., & Somers. K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, *49*, 974-997.
Pérez, E., & Medrano, L. (2010). Análisis factorial exploratorio: bases conceptuales y metodológicas [Exploratory factor analysis: Conceptual and methodological foundations]. *Revista Argentina de Ciencias del Comportamiento, 2*(1), 58-66.
Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*, 13-32.
Revelle, W. (in press). *An introduction to psychometric theory with applications in R.* Springer.
Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, *45*, 73-103.
Satorra, A., & Bentler, E. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye and C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
Seldmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309-316.
Steiger, J. H., & Lind, J. M. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.
Tanaka, J. S., & Huba, G. H. (1989). A general coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *42*, 233-239.
Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.
Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement*, *45*, 203-209.
Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: The University Chicago Press.
Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321-327.
Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin and E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy*. Norwell, MA: Kluwer Academic.
Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*(2), 231-251.
Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*(1), 1-28.
Vigil- Colet, A., Canals, J., Cosí, S., Lorenzo-Seva, U., Ferrando, P. J., Hernández-Martínez C., … Domenech, E. (2009). The factorial structure of the 41-item version of the screen for child anxiety related emotional disorders (SCARED) in a Spanish population of 8 to 12 years-old. *International Journal of Clinical and Health Psychology*, *9*(2), 313-327.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables. Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56-75). Newbury Park, CA: Sage.
Williams, B., Brown, T., & Onsman, A. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*(3), 1-13.
Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany: State University of New York Press.
Yuan, K.H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, *51*, 289-309.
Zwick, R.W. y Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442.

(Article received: 8-2-2014; revision received: 27-5-2014; accepted: 28-5-2014)