anales de psicología, 2014, vol. 30, nº 3 (octubre), 1170-1175
http://dx.doi.org/10.6018/analesps.30.3.199991

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

Exploratory item factor analysis: Additional considerations

Pere J. Ferrando*, and Urbano Lorenzo-Seva

CRAMC (Centro de Investigación para la Medida de la Conducta); Departamento de Psicología; Universitat Rovira i Virgili (Tarragona, Spain)

Título: El análisis factorial exploratorio de los ítems: algunas consideraciones adicionales.
Resumen: El presente artículo puede considerarse como una ampliación al trabajo de Lloret et al. (2014) en la que se discuten, de forma ampliada, dos tópicos de especial relevancia en análisis factorial de ítems: (a) la decisión acerca de la matriz de correlación más apropiada en cada caso, y (b) la determinación de soluciones finales semi-confirmatorias, que sean realistas, interpretables y que utilicen la información disponible por el investigador. La presentación de los dos tópicos no es neutral, sino que refleja las posiciones de los autores, por lo que debe ser también evaluada críticamente por parte del lector. En ambos casos se ofrecen recomendaciones prácticas. Sin embargo, el trabajo va especialmente dirigido a los investigadores aplicados con cierta orientación metodológica que quieren ir un poco más allá de las recomendaciones actualizadas propuestas anteriormente.
Palabras clave: correlaciones producto-momento; correlaciones policóricas; modelo de respuesta graduada; estructura simple; rotaciones semi-especificadas; matriz diana; rotaciones ortogonales y oblicuas.

Abstract: This article can be considered to be an extension of a previous work by Lloret et al. (2014) in which we discuss in more depth two particularly relevant topics in item factor analysis: (a) how to chose the most appropriate correlation matrix, and (b) how to arrive at a realistic and interpretable semi-confirmatory solution by making use of the available information. The discussion of these topics strongly reflects the views of the authors. So, we encourage the interested reader to take it critically. For both topics we offer practical recommendations. However, the article is mainly intended for applied researchers with a certain methodological penchant and who want to go beyond standard recommendations.
Key words: Product-moment correlations; Polychoric correlations; Graded Response Model; Simple Structure; Semi-Specified rotations; Target Matrix; Orthogonal and Oblique Rotations.

Few psychometric techniques have generated as much controversy and had as many ups and downs as exploratory factor analysis (EFA). Regarding the first point, it has been and is still by far the most used technique in item analysis, but also the most criticized. As for the second point, the increasingly widespread use of structural equation models has consigned EFA for many years to the limbo of almost obsolete techniques. At most, an EFA might have been justified in the early phases of a study. However, from here, the 'correct' technique was the more rigorous confirmatory factor analysis (CFA). Reality, however, is obdurate: the structure of most tests is inherently complex and does not fit well with the restrictive CFA hypotheses. It appears that EFA is still needed to analyze item responses.

Psychometricians, for their part, have not abandoned researching in EFA and making it evolve. However, due to lack of interest by some and inertia by others, these evolutions have not been seen in the applied field. The authors, therefore, unreservedly applaud both the initiatives of the editor of Annals of Psychology and the work of Lloret et al. (2014) that precedes ours. It is a current, correct, clear and didactic guide on how the applied researcher must proceeds using EFA in the analysis of items and tests.

In this work we try to discuss two topics that we believe are of central interest and which, due to space restrictions, were not fully dealt with in the previous article: the choice of the most appropriate correlation matrix in each case (or rather, as will be seen, of the most appropriate FA model), and that of the intermediate solutions between EFA and CFA, an issue which in our view will become increasingly important in the applied field. The treatment is still aimed at the FA user but is above all oriented toward those researchers interested in methodology who intend to go a step further in the knowledge of evolutions of the technique and the lines that will receive more interest in the future.

*** Dirección para correspondencia [Correspondence address]:**
Pere J. Ferrando, Departamento de Psicología; Universitat Rovira i Virgili; Carretera de Valls s/n; 43007 Tarragona (Spain).
E-mail: perejoan.ferrando@urv.cat

Product-moment correlations or polychoric correlations?

The title of the section intentionally reflects how this question has traditionally been raised. It is about trying to decide the most appropriate association matrix to estimate the linear or common FA model. However, in taking this decision, we are actually deciding between two different factor models: linear and nonlinear. More specifically, when an FA is fitted to a polychoric correlation matrix what is actually being fitted is a nonlinear model of item response theory: Samejima graded response model (1969, see Ferrando & Lorenzo-Seva, 2013).

In the unidimensional case, which is the clearest from an illustrative viewpoint, the linear model (i.e., FA on Pearson) assumes that the regression of the item score on the common factor is linear and with constant variance. The nonlinear model (i.e., FA based on polychorics) instead assumes that such regression is an S-shaped curve (ogive) and that variance decreases toward the extremes. Given that item responses are limited (e.g., between 0 and 1 or 1 and 5) and the factor is conceived as an unlimited variable, the nonlinear model is more plausible and theoretically more appropriate. The linear model, therefore, should always be seen as an approximation.

The first point to discuss refers to the determinants that make the linear approximation good or not (assuming that the nonlinear model is correct). The primaries are two: (a) the position or difficulty of the items (or, equally, how extreme item distributions are) and (b) their discriminating power. When items are of medium difficulty and moderate discrimination, regression curves have a moderate slope and are centered on the mean of the factor distribution. Thus, regression is essentially linear for most responding subjects and the Pearson-based FA model is a good approximation. At the opposite end, the linear approximation will be poor when items are both extreme and highly discriminating. The number of response categories also plays a role, but is not the main determinant as commonly believed. As the number of categories rises, distributions tend to become less extreme and the attenuation problems discussed below are minimized. However, linear FA is sometimes a good approximation for binary items and a poor approximation for continuous items (Ferrando, 1994; McDonald & Ahlawat, 1974). Finally, from a practical standpoint, we recommend evaluating the two primary determinants through the inspection of the means and the skewness coefficients of the items (position) and the magnitude of the inter-item correlations (discrimination). If means are not too extreme, the skewness coefficients are not greater than 1 in absolute value, and the inter-item correlations move, perhaps around .50 or less, the linear model is expected to function reasonably well. These recommendations are consistent with the results obtained by Muthén and Kaplan (1985) using simulation.

If the linear model is fit under inappropriate conditions, the main consequences expected are twofold: (a) spurious evidence of multidimensionality and (b) differential attenuation of the factor loadings (Ferrando & Lorenzo-Seva, 2013). As for the first point, if a curve is to be fit by a linear model, additional terms or curvature factors will be needed to obtain a good fit. These factors are traditionally known as 'difficulty factors' (Ferrando, 1994; McDonald & Ahlawat, 1974) and have no substantive interpretation. Contrarily, as the reader may surmise, its composition will be related to the difficulty of items and their discriminating power. In summary, if a model with an adequate number of substantive factors is assessed, this model will possibly provide insufficient fit, and it will be necessary to estimate additional factors that are not interpretable.

As for the second point, the loading estimates provided by the linear model will be biased downwards (i.e. attenuated) regarding the true loadings and, in addition this effect will not be constant: attenuation will be greater the more extreme and discriminatory the items (Muthén & Kaplan, 1985). Since the factor loading is the main indicator of item quality as a measure (Ferrando, 2009), this problem has important consequences as it may lead to incorrect interpretations.

At this point, the reader might ask the following: Although linear FA is a good approximation under certain conditions, why not always use the most correct nonlinear model? Well, in general terms, because it is not unusual for an approximate but simple and robust model to work better than a theoretically more appropriate but more complex and unstable one, and this is quite common in the case of FA, (Ferrando, 2009; Ferrando & Lorenzo-Seva, 2013, Hofstee, ten Berge, & Hendricks, 1998). Nonlinear FA is, of course, rather more complex than linear FA and presents potential problems that we should be aware of and which will be discussed later.

The first group of problems refers to the precision and stability of the polychoric correlations serving as a basis for non-linear FA. The polychoric correlation is not a statistic obtained directly from the data, but an estimator of a latent correlation between assumed continuous response variables that is estimated iteratively and which is quite complex (e.g., Rigdon, 2010). This estimator may not converge, achieve implausible results, or may simply be very imprecise, with typical errors much larger than those of a Pearson correlation. Among other factors, these problems depend on the sample size and the number of response categories. The impact of these factors can be seen by bearing in mind that the polychoric correlation is calculated from the contingency table between the scores of two items and that to obtain stable estimates cells must contain a reasonable number of cases (Mislevy, 1986). The more response categories there are, the greater the number of cells and therefore the more sample potentially needed to fill them. We agree that with samples smaller than 200 cases the nonlinear model is not advisable. Nevertheless, larger samples alone do not guarantee stability of the inter-item correlation matrix, which is the basis for the analysis. It is worth remembering that if the underlying correlations are not stable; the estimated loadings will be even less stable when these correlations are factored.

The origin of the second set of problems is that each polychoric correlation is an independent maximum-likelihood estimator of an assumed correlation between two latent variables but the polychoric correlation matrix as a whole is not a consistent estimator of a population correlation matrix. Even if all estimates are plausible, some will have more error than others. The most appropriate simile could be a Pearson correlation matrix where each correlation would have been obtained from a different sample size: the estimated correlations in large samples would possibly be more correct and more reliable than those obtained in small samples. This situation can lead to several problems. From the outset, the polychoric matrix may not be positive definite and therefore some FA procedures (in particular maximum likelihood) are simply inapplicable. Although applicable, certain estimates will not generally be appropriate. More specifically, the goodness of fit indicators derived from the chi-square statistic will be incorrect and generally inflated. For this reason, we advise estimating the FA of polychoric correlations using ULS, and evaluate fit through indicators that do not depend directly on the chi-square: the GFI index and the RMSR (e.g. McDonald, 1999).

The ULS parameter estimates are consistent (Mislevy, 1986); therefore, under reasonable conditions, the FA of the polychoric matrix is expected to reach correct estimates, and simulation studies generally agree. If data are generated from a non-linear model, then the ULS solution based on the polychoric matrix recovers the true factor loadings better than the linear ULS solution (e.g., Lee, Zhang, & Edwards, 2012).

This is to be expected as the first procedure corrects the differential attenuation problem discussed above. The assessment of fit, however, is another subject. Our experience with real data indicates that the adjustment provided by the polychoric FA is generally worse than that obtained based on Pearson's correlations, even when conditions suggest it should be better. Moreover, this result becomes more apparent the greater the number of items is (see also, Rigdon & Ferguson, 1991). In summary, in conditions where it is clearly more appropriate a priori, nonlinear FA generally leads to more accurate estimates of factor loadings than linear FA. However, the correct assessment of fit remains a problem. This problem could perhaps be improved by using weighted least squares procedures, such as those proposed by Muthén (1993) for the confirmatory case, which take into account the different degrees of error of the polychoric correlations.

In conclusion, the problem of deciding the most appropriate FA approach to evaluate responses to a test is complex, and we accept that our discussion of the subject may have caused confusion. It would be easier to provide a series of simplistic or categorical recommendations, but this would be deceiving the reader, and it is precisely these stereotyped recommendations that the article by Lloret et al. (2014) rightly criticizes. As a guide, we propose that the researcher considers or evaluates the following aspects: (a) position and distribution of responses, (b) magnitude of inter-item correlations, (c) number of response categories, and sample size. In some cases this review will lead to clear decisions. Thus, a questionnaire consisting of 7-point Likert-format items that have medium difficulties and are not excessively discriminative, and which is administered to a sample of 250 subjects justifies the use of linear FA based on Pearson correlations. This scenario is relatively common in measures of personality or attitude that evaluate non-pathological traits (Ferrando, 2009). At the other extreme, a test containing very easy items along with very difficult items, all very discriminative, with a 3-point response format and administered to a sample of 600 subjects clearly suggests the use of the non-linear option. This second scenario is not so usual, but given in some ability tests and clinical questionnaires (Reise and Waller, 2009). In the intermediate field where most cases will be located, our recommendation is to carry out the two types of analysis and evaluate both solutions. In a program like FACTOR (Lorenzo-Seva and Ferrando, 2013), just a click of the mouse is needed to move from one option to another, and the additional information obtained is always worthwhile.

To conclude this section, we would like to mention some actions that could greatly improve the non-linear option. First, it would be useful to implement more robust and, if possible, non-iterative estimation methods for polychoric correlations. Until now, the procedure that has worked best is the Bayesian estimation based on the MAP criterion and with a moderately restrictive prior distribution. This method ensures convergence for all the correlation estimates and also leads to plausible estimates in all, even in small samples. Second, the standard errors of polychoric correlations would have to be provided in order for the researcher to evaluate their stability and accuracy before subjecting them to FA. Finally, it would be worthwhile studying and eventually implementing FA estimation methods based on weighted least squares, which might enable us to better evaluate the fit of the proposed solution. The authors are working on these points in order to implement them in the FACTOR program at some point in the future (Lorenzo-Seva and Ferrando, 2013).

Intermediate factor solutions between EFA and CFA

As already stated, the appearance of CFA relegated the use of EFA to a *minor* technique justified only in the initial stages of research. The juxtaposition between the two approaches was such that they were even understood as two completely different techniques. Today there is unanimity in that, as Lloret et al. (2014) expose, EFA and CFA are two ends of a continuum. It should be noted that, despite their names, how both techniques are commonly used in applied research makes neither one fully *exploratory* nor the other fully *confirmatory*. The reason is explained below.

The usual application of EFA involves computing an objective procedure that enables deciding the optimal number of factors to be extracted, followed by a factorial rotation that maximizes the criterion of *simple structure*. This application ensures EFA is not merely an exploratory technique: (1) the number of dimensions is usually decided using procedures that favor models with few factors (which indicates a type of factor model that is favored); and (2) the simple structure criterion is in itself a model of factor solution (a model that expects a simple pattern of factor loadings). While it is true that the simple structure criterion as defined by Thurstone (1947, p. 335) is impossible to fit, Kaiser (1974) redefined it as *factorial simplicity*: a model that expects each variable to show only a single substantial loading different from zero in a single factor. Since an implicit model is being used, the final solution is not purely *exploratory*.

At the other extreme, CFA explicitly implies proposing a population model that determines the number of factors, as well as the loadings that each variable must show in each factor. It is noteworthy that it is usual for researchers to define for each variable what loadings should be zero in the population, giving freedom of magnitude to the one expected to be different from zero. However, in the practical application of CFA it is quite common that once an initial model is fit using the data of a particular sample, the researchers relax *ad hoc* some of the model parameters (for example, allowing in the model that some of the variables show more than one loading other than zero) to favor a better model fit. When the parameters of the initial model are thus relaxed, the final solution is no longer merely *confirmatory*.

Now, if the EFA and CFA are two ends of a continuum, the natural question is: what methodological options does the researcher have available between either end? In the following text we present the options available starting from the most confirmatory options and moving increasingly toward more exploratory options.

As already stated, the structure of most psychological tests is complex and difficult to model in terms of CFA's restrictive hypotheses. Perhaps the most restrictive hypothesis of factor models is to assume that in the population each item is a pure indicator of a single underlying dimension: i.e., to propose that each item will present a single factor loading with absolute value 1 and all other loadings will be exactly zero. A first available technique to relax this restrictive model is the *Procrustean rotation* (referring to the bandit from Greek mythology who made his victims fit his bed either by stretching their limbs or cutting them off). Orthogonal Procrustes rotation was proposed by Cliff (1977), while oblique rotation was proposed by ten Berge and Nevels (1977). The idea of this rotation is to propose a *target* matrix with values 1 and 0 indicating the expected factor pattern for each item. For example, if we imagine a set of 20 test items, from which it is hypothesized that 10 items measure Extraversion and another 10 measure Responsibility. The target matrix will consist of a matrix of 2 columns (one column per factor) and 20 rows (one for each item), and will only contain values of ones and zeros. The 10 items expected to be related to Extraversion would present a value of 1 in a column in the target matrix (representing the Extraversion factor), and zero in the other column (representing the Responsibility factor). The expected factor loading pattern of the 10 Responsibility items will be equally represented (in this case with the pattern of ones and zeroes according to the hypothesized Responsibility factor). Once the target matrix is proposed, the rotation seeks the position that minimizes the distance between the loadings in the rotated factorial pattern with respect to the target matrix. It should be noted that even in a good fit of the model; the rotated factor pattern does not show exact values either of ones or of zeros, but rather values close to ones and zeros. To assess to what extent the rotated solution is congruent with the target matrix, it is customary to calculate the coefficient of congruence (Tucker, 1951): a value between .85 and .94 indicates acceptable factorial congruence, while values above .94 indicate good factorial congruence (Lorenzo-Seva & ten Berge, 2006). This type of rotation is known as a *fully specified Procrustean rotation* (since the target matrix hypothesizes how *all* loadings of the rotated factor pattern should be).

The next step toward the relaxation of the constraints of the model is known as a *semi-specified Procrustean rotation* and was proposed by Browne in both orthogonal rotation (Browne, 1972a) and oblique rotation (Browne, 1972b). The aspect of the model that is relaxed is to suppose that each item must present a perfect loading (i.e. a value of 1) in the factor with which the item is related. The aspect of the model that does *not* relax is to assume that all items are good indicators of either factor (and therefore show loadings close to zero in the factors with which they are not related). Once more, a target matrix is proposed, although only the value of some of the elements of the target matrix is specified, leaving the remainder unspecified. The target matrix values that are specified correspond to the values expected to be zero in the rotated factor pattern. Following our example of 20 items, the target matrix will equally consist of a matrix of 2 columns (one column per factor) and 20 rows (one for each item). The 10 items expected to be related to Extraversion will present in the target matrix a free value in one column (representing the Extraversion factor), and zero in the other column (representing the Responsibility factor). The expected factor loading pattern of the 10 Responsibility items will also be equally represented (in this case with the pattern of free values and zeros according to the hypothesized Responsibility factor).

Once again, it is worth noting that it is most likely that even in a good model fit, the rotated factor pattern does not show exact values of either one or zero, but rather values close to zero (for specified loadings) or markedly different of zero (for unspecified loadings). The semi-specified procrustean rotation is less restrictive as it is not hypothesized that all items are *pure indicators* of one factor or another: rather, all items are considered to be *good markers* of either test factor.

The next step toward relaxation of model constraints is known as fitting an independent-cluster-basis solution (McDonald, 2000, 2005). Lloret et al. (2014) state it is a question of identifying the factors rotated according to a reduced number of good markers (between two or three markers per factor). From a computational viewpoint, the fit to this model can be obtained through a semi-specified rotation in which only a few rows of the target matrix have zero values. In our 20-item example, 2 extraversion items would be selected as the extraversion factor markers, while 2 other items would be selected as Responsibility factor markers. Therefore the target matrix would only specify the values of these 4 items expected to be close to zero in the rotated pattern. The particular position of the zeros in the target matrix will depend on the hypothesized 4 items that are defined as markers of the factor. On the other hand, the other 16 items will not show any specification in the target matrix. The fit of independent clusters is less restrictive by the simple fact that it is only hypothesized that some items are good markers of one or another test factor.

So far, the relaxation of the model has passed through defining a target matrix on which we have been successively reducing the number of restrictions. However, the researcher has had to propose at least a certain number of markers per factor. The next step towards relaxation is to avoid having to propose the markers for each factor. This type of fit is achieved by rotation procedures that build a target matrix themselves, and assign them the values to be specified. An example of this type of rotation is Promin (Lorenzo-Seva, 1999). Promin is based on: (1) using weights proposed by Cureton and Mulaik (1975) to identify potentially simpler items (i.e., the best markers) before starting rotation so that it is precisely those items that guide the rotation; (2) construct a semi-specified target matrix (where values expected to be close to zero are specified); and (3) compute the oblique semi-specified procrustean rotation. We might say that Promin takes the difficult decisions in the proposal of the model (from the degree of specification of the model, to the particular values specified). In this way, Promin is very close to the purely exploratory end of the continuum as far as the researcher is concerned (who will not advance any parameter of the model except the number of dimensions).

Given the exploratory and confirmatory continuum just described, we wonder where that leaves the classical rotation procedures (such as Varimax, Oblimin or Promax). These rotation methods aim to obtain a factorial pattern as close as possible to the criterion of *factorial simplicity*. To perfectly comply with this criterion, it would be necessary for all items to be perfect indicators of one factor or another (although it is true that the researcher does not have to decide for any item what factor would be an indicator of the item), resulting in a factorial pattern rotated with exact values of only ones or zeros. Although it is not easy to decide where these procedures should be placed in the continuum we have described, it appears clear that they propose a model with constraints (factorial simplicity itself for all items without exception) away from the purely exploratory end. Therefore, applying, for instance, Oblimin to rotate a factorial solution where some items are not pure indicators of one factor or another (i.e. items that are complex in nature), has the negative consequence that its complexity is distributed to a greater or lesser extent to all rotated pattern loadings (due to attempting to minimize the complex structure of some items). In other words, the rotation procedure would not allow us to correctly identify the degree to which some items are good markers of some factors.

Finally, we need to find a position on a commonly used restriction in the context of the EFA psychological tests: the orthogonality of the rotated factors. Izquierdo et al. (in press) find that a high proportion of studies in the context of the EFA use the Varimax rotation procedure (i.e., one that imposes orthogonality of factors).In our opinion, imposing the orthogonality of the rotated factors prevents detecting the possible dependence relation between factors. On the other hand, when the obliquity of the factors is allowed, orthogonality is not being imposed: i.e., if an oblique rotation procedure is used, but the factors are actually independent of each other, the rotated factor solution will show correlations between factors close to zero (and therefore negligible). Therefore, our advice is to systematically apply oblique rotations, as advised by Browne in his revision of the factorial rotation procedures of 2001.

Our FACTOR program (Lorenzo-Seva and Ferrando, 2013) incorporates all the unrestricted rotation procedures discussed in this paper. The responsibility of the applied researcher is to decide according to the psychological test they wish to analyze where they stand in the continuum between the exploratory confirmatory ends of the continuum.

Discussion

In recent years, and especially here in Spain, a series of updated EFA guidelines and reviews have been published aimed at applied researchers. It seems clear that the technique is still alive and still arouses interest. If, moreover, (as we hope) these recommendations end up in the applied field, the level of research that uses EFA in Spain (which is already substantial) will greatly improve.

The central discourse of our contribution, however, is that another step can be taken. The researcher inclined to do so can look further into the bases and the logic of some key points of EFA, and this acquired knowledge, in turn, will allow the best decisions to be taken in each case. Decisions that, as we have seen, go beyond general recommendations, require thorough evaluation by the researcher, and refer to key points such as choosing the most appropriate model - linear or non-linear- given the characteristics of their data or the transformation most in line with the hypothesis of their research and the information available.

Acknowledgements.- The research was partially supported by grants from the Catalan Ministry of Universities, Research and the Information Society (2009SGR1549) and from the Spanish Ministry of Education and Science (PSI2011-22683).

Referencias

Browne, M. W. (1972a). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 115-120.

Browne, M. W. (1972b). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 207-212.

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

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, *31*, 33-42.

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

Ferrando, P. J. (1994). El problema del factor de dificultad: una revisión y algunas consideraciones prácticas.*Psicológica.* 15(2), 275-283.

Ferrando, P. J. (2009). Difficulty, discrimination and information indices in the linear factor-analytic model for continuous responses. *Applied Psychological Measurement*, 33, 9-24.

Ferrando, P. J. y Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory. Technical Report*. Department of Psychology, Universitat Rovira i Virgili. Retrieved from http://psico.fcep.urv.cat/utilitats/factor/

Hofstee, W. K. B., Ten Berge, J. M. F., y Hendricks, A. A. J. (1998). How to score questionnaires. *Personality and Individual Differences, 25*, 897-910.

Izquierdo, I., Olea, J. y Abad, F.J. (en prensa). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema.*

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36.

Lee, C. H., 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.

Lloret, S., Ferreres, A., Hernández, A. y Tomás, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología, 30* (3)*,* 1151-1169.

Lorenzo-Seva, U. (1999). Promin: a method for oblique factor rotation. *Multivariate Behavioral Research*, *34*, 347-356.

Lorenzo-Seva, U.y Ferrando, P. J. (2013). FACTOR 9.2 A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied Psychological Measurement*, *37*, 497-498.

Lorenzo-Seva, U.y ten Berge, J. M. F. (2006). Tucker's Congruence Coefficient as a Meaningful Index of Factor Similarity. *Methodology*, *2*, 57-64.

McDonald, R. P. (1999). *Test theory: a unified treatment*. 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. y Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*. 27, 82-99.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of educational statistics, 11*(1), 3-31.

Muthén, B. (1993). Goodness of fit with categorical and other nonnormal variables. En K.A. Bollen y J.S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park: Sage.

Muthén, B. y 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.

Reise, S. P. y Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5,* 27-48.

Rigdon, E. E. (2010). Polychoric correlation coefficient. En N. Salkind (Ed.), *Encyclopedia of research design*. (pp. 1046-1049). Thousand Oaks, CA: SAGE Publications, Inc.

Rigdon, E. E. y Ferguson, C. E. (1991) The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research, 28*, 491-497.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17*. Iowa City: Psychometric Society.

Ten Berge, J.M.F., y Nevels, K. (1977). A general solution to Mosier's oblique Procrustes problem. *Psychometrika*, *42*, 593-600.

Thurstone, L.L. (1947). *Multiple Factor Analysis*. Chicago: University of Chicago Press.

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. *Personnel Research Section Report No.984*, Washington D.C.: Department of the Army.

(Artículo recibido: 27-5-2014; revisión recibida: 5-6-2014; aceptado: 8-6-2014)