Papeles del Psicólogo, 2010. Vol. 31(1), pp. 18-33
http://www.cop.es/papeles

FACTOR ANALYSIS AS A RESEARCH TECHNIQUE IN PSYCHOLOGY

Pere Joan Ferrando and Cristina Anguiano-Carrasco
Research centre for the assessment and measurement of behaviour, Rovira i Virgili University, Tarragona

The present article explains the main aspects of factor analysis as a tool in psychological research. First of all, the basic issues are reviewed at a conceptual level, so that the review is appropriate for beginners as well as for those seeking more in-depth knowledge of the technique. Next, the differences between exploratory and confirmatory analysis are discussed in some detail, as well as the procedures for fitting the model and transforming the initial solution. These aspects are discussed in accordance with the recommended steps in factor-analytic research: from the design and data collection to the interpretation of the final solution. Finally, the functioning of the "Factor" program is explained. Factor is more complete than most commercial programs, and is also freely distributed. Additionally, we provide readers with an exercise (and its solution) for practicing the application of the material discussed in the text.
Key words: *Exploratory factor analysis, Confirmatory factor analysis, Principal components, FACTOR computer program.*

El presente texto explica los principales aspectos del análisis factorial como instrumento de investigación psicológica. Se revisan en primer lugar los aspectos básicos a nivel conceptual, de modo que su lectura sea adecuada tanto para el lector principiante como para aquellos que quieran profundizar más en sus conocimientos de la técnica. Después, se discuten con cierto detalle las diferencias entre el análisis exploratorio y confirmatorio, y los procedimientos para estimar el modelo y obtener la solución transformada. Estos puntos se discuten siguiendo cada una de las etapas recomendadas en una investigación: desde el diseño y recogida de datos hasta la interpretación de la solución final. Finalmente, se explica el funcionamiento del programa Factor, un programa más completo que la mayor parte de los programas comerciales, y que además es de distribución libre. Adicionalmente, se propone al lector la realización de un ejercicio resuelto, a modo de práctica para la aplicación de lo expuesto en el texto.
Palabras clave: *Análisis factorial exploratorio, Análisis factorial confirmatorio, Componentes principales, Programa de ordenador FACTOR.*

Since its inception at the beginning of the 20th century, factor analysis (FA) has undergone considerable developments. The simple initial model proposed by Spearman (1904) for validating his theory of intelligence has given rise to an extensive family of models for use not only in the social sciences but also in other domains, such as Biology or Economics. Given that a thorough treatment of FA would go well beyond the possibilities of the present article, it makes sense to first of all specify the aspects that we shall cover here.

For many years now, the first author has been reviewing empirical studies that use FA in psychological research, and the experience acquired will help to make the initial selection of issues on which to focus. First of all, most factorial studies in psychology use FA to assess (a) the structure of a test based on the scores on its items, or (b) dimensional hypotheses using as measures the scores on different tests. It seems reasonable, then, to focus this exposition on this type of measure: scores on items or tests.

Secondly, experience shows that the methodological problems involved in such studies tend to be similar. A first group of problems emerges at the design stage of the research (a stage usually somewhat neglected in factorial studies). The problems in the second group have to do with the decisions the researcher must make at the stage of parameter estimation and assessment of model-data fit and at that of rotation. In particular, the majority of the problems arise from the unjustified use of the package known as "Little Jiffy", which involves (a) Principal component (b) eigenvalues-greater-than-one rule, and (c) Varimax rotation. We shall devote particular attention to design, parameter estimation and assessment of model-data fit issues.

Even with such delimitations, the issue remains too broad. In the present article we shall concentrate on the most basic general model of FA: the linear model, based on correlations, and which analyzes measures obtained in a single group of participants and on a single occasion.

Correspondence: Pere Joan Ferrando. *Universidad 'Rovira i Virgili'. Facultad de Psicologia. Carretera Valls s/n. 43007 Tarragona. España. E-mail: perejoan.ferrando@.urv.cat*

This limitation means leaving out topics of great interest: extended models of means and covariances, non-linear models and their relations with item response theory, and models for multiple groups and multiple occasions. Nor, indeed, shall we be able to deal with the general issue of factor scores.

The article's approach is conceptual and applied, and we have attempted as far as possible to avoid formal expressions and technical detail. Only the basic equations of the model are included, and these are presented in separate boxes. Likewise, we have tried to minimize the use of references, at the same time offering a section on recommended reading. In this regard, we should stress that some of the topics discussed are controversial, and that the recommendations we make in the article reflect our own theoretical positions. The recommended reading section gives readers the opportunity to consider other positions and to critically evaluate what we have set out here.

THE BASIC IDEAS OF FACTOR ANALYSIS

FA is a statistical model that represents the relations among a set of variables. It is based on the assumption that these relations can be explained from a series of non-observable (latent) variables called factors, the number of factors being substantially smaller than the number of variables. The model is obtained directly as an extension of some of the basic ideas of linear regression and partial correlation models. Linear regression provides the basic FA equations, whereas partial correlation theory provides the key ideas for assessing model-data fit.

In the linear regression model, the score on a criterion variable can be explained in part by a linear combination (a sum of variables, each one of which is multiplied by a weight or coefficient) of a series of predictor or explanatory variables called regressors. It is explicitly assumed that the combination is no more than an approximation, and that a part of the criterion score will not be predictable or explainable from the regressors. This unexplained part is the error term (see eq. 1 in the box).

$$Y = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m + e \qquad (1)$$

$$X_j = \mu_j + \lambda_{j1}f_1 + \lambda_{j2}f_2 + \ldots + \lambda_{jm}f_m + e_j \qquad (2)$$

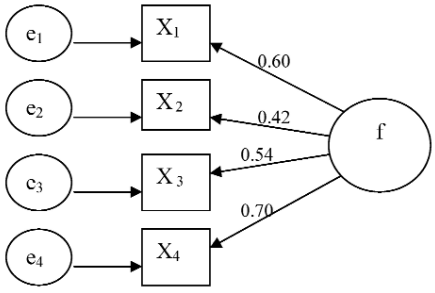
$$X_j = \mu_j + \lambda_j f + e_j \qquad (3)$$

In FA a set of observable variables (items, subtests or tests) is analyzed, each one of which can be considered as a criterion. If this conceptualization is used, FA would consist in a system of regression equations such as that described above (one equation for each observable variable) in which the regressors, here called factors, are common for a subset (common factors) or the whole set (general factors) of variables (see eq. 2 in the box). For each one of these equations, the basic difference between FA and a conventional regression is that the regressors, that is, the factors, are not observable. This difference is what makes FA a more complex model than that of regression. For a start, since the factors are unobservable, they lack a particular measurement scale. To solve this, the simplest practice, which we shall employ here, consists of assuming that the factors are on a standard scale: with zero mean and unit variance. If, in addition, the observable variables are also standardized, the model is mathematically simpler and easier to interpret.

By analogy with the regression model, it follows that the simplest FA model is that which proposes a single general factor (eq. 3). This model would be equivalent to that of simple regression, and was the initial FA model set out by Spearman. In order to get started and clarify some ideas we shall study a Spearman solution based on an FA of a 4-item set, together with its representation in a Wright Diagram. As regards the drawing up of such diagrams, we refer the reader to the article by Ruíz, Pardo and San Martín in the present issue of this journal.

Ítem	f_i
X_1	0.60
X_2	0.42
X_3	0.54
X_4	0.70

Represented graphically this would be:



The FA equation for item 1 is:

$$X_{i1} = 0.6\, f_i + e_{i1}$$

, which is interpreted as follows. The score of individual *i* on item 1 is partly determined by the effect of the common factor (the level of *i* in the common factor *f*) and is partly error. In FA the error term includes all those effects other than the common factor or factors that influence the score on the observed variable. More systematically, we can distinguish three broad groups of effects or types of error: (a) sampling error (statistical error), (b) measurement error (psychometric error) and (c) error of approximation. This last error component means that the specified model is not considered exactly correct even in the population. Indeed, models are not, nor pretend to be, exact: they are, at best, reasonable approximations to reality.

The value 0.6 is the factorial weight, and is equivalent to the slope in the regression model. If item and factor scores are on a standard scale, this weight reflects the importance of the factor in the determination of the score on this item. The greater the weight, the greater the importance of the factor, and therefore, the less the influence of the error. Moreover, on the assumed standard scale, this weight can be interpreted as the correlation between the factor and the item. Its square, which is the coefficient of determination, is interpreted as the proportion of variance in the scores for this item that can be explained from the factor. Thus, in accordance with our estimated solution, item 1 would correlate 0.6 with the general factor; this factor would explain 36% of the variance in the scores for this item (0.62=0.36), so that the remaining 64% would be error variance. In the terminology of FA, the proportion of explained variance is referred to as "communality".

Before continuing, it may be an interesting exercise for the reader to interpret the FA equation corresponding to another item, say, item 2. Likewise, it is very important to take into account that the above interpretations are only valid if the variables and the factor are on a standard scale. Should this not be the case, the weight does not have a clear interpretation, since it would reflect to a greater or lesser extent the differences in the measurement scale of the variables. When interpreting the output of an FA it is essential to ensure that the solution interpreted is a standardized solution.

When more than one factor is hypothesized we have the multiple FA model, also known as Thurstone's (1947) model, after its principal developer. The model is the same for any number of factors and, for simplicity's sake, we shall explain it with two. The key aspect here is the hypothesized relation between the factors. As in regression, the simplest and most easily interpretable case is that in which the factors are uncorrelated (conceptually, they are independent of one another). A solution of this type is called an "orthogonal solution". In an orthogonal solution the factorial weights continue to be interpreted as variable-factor correlations, their squares are proportions of variance explained by the corresponding factor, and the sum of these squares is the communality (see eqs. 4 and 5), or the proportion of variance explained jointly by the factors

$$\sigma_j^2 = \lambda_{j1}^2 + \lambda_{j2}^2 + \sigma_{ej}^2 = 1 \qquad (4)$$

$$1 = h_j^2 + \sigma_{ej}^2 \qquad (5)$$

$$r_{jk} = \lambda_{j1}\lambda_{k1} + \lambda_{j2}\lambda_{k2} \qquad (6)$$

The case of correlated factors, called 'oblique solution' is the most complex, but also perhaps the most realistic in practice. The most important aspect on interpreting a solution of this nature is that now the weights and the variable-factor correlations are different coefficients. As in the theory of linear regression, the factorial weights are now standardized regression coefficients and measure the effect of the factor on the response variable when the other factors remain constant. These weights are presented in the matrix called "factor pattern". On the other hand, the variable-factor correlations are called structural coefficients, and are presented in the matrix called "factor structure". The equations corresponding to the weights and structural coefficients are presented in the box below. Conceptually, the weights indicate the extent to which the factor influences the variable, whilst the structural coefficients indicate the degree of similarity between factor and variable. In the case of oblique solutions, we shall focus above all on the weights, that is, the pattern matrix.

$$X_{ij} = \lambda_{j1} f_{i1} + \lambda_{j2} f_{i2} + \varepsilon_{ij} \qquad (7)$$

$$s_{j1} = \lambda_{j1} + \lambda_{j2}\phi_{12} \qquad (8)$$

$$\sigma_j^2 = \lambda_{j1}^2 + \lambda_{j2}^2 + 2\lambda_{j1}\lambda_{j2}\phi_{12} + \sigma_{ej}^2 \qquad (9)$$

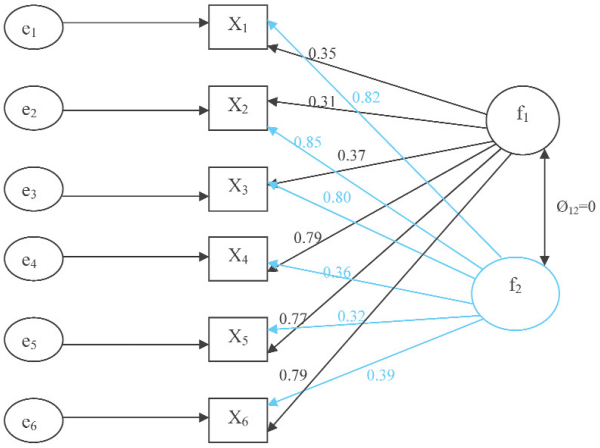
$$r_{jk} = \lambda_{j1}\lambda_{k1} + \lambda_{j2}\lambda_{k2} + \phi_{12}\left(\lambda_{j1}\lambda_{k2} + \lambda_{j2}\lambda_{k1}\right) \qquad (10)$$

In equation (8) s_{j1} is the structural coefficient (variable-factor correlation) and ø is the correlation between factors.

Let us look now at an example of a multiple orthogonal solution with 6 items and two factors:

Item	F_1	F_2
X_1	0.35	**0.82**
X_2	0.31	**0.85**
X_3	0.37	**0.80**
X_4	**0.79**	0.36
X_5	**0.77**	0.32
X_6	**0.79**	0.39

The graphical representation is:



Wright Diagram for two uncorrelated common factors.

Taking as an example item 2, we obtain the following equation:

$$X_{i2} = 0.31 f_1 + 0.85\, f_{i2} + e_{i2}$$

This equation indicates that individual *i*'s score on item 2 is determined not only by the error but also, and primarily, by the second factor (0.85), and to a lesser extent by the first factor (0.31). In this case, the communality or explained variance is obtained by adding the squares of the weights on both factors, so that for item 2,

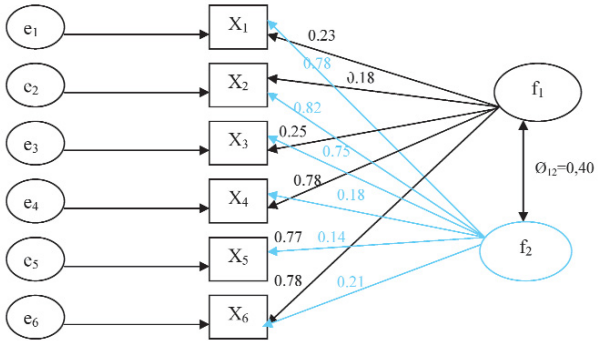
$$h^2_2 = 0.31^2 + 0.85^2 = 0.82$$

Subtracting the communality from one we obtain an error variance of 0.18. In terms of proportions, the two factors jointly explain 82% of the total variance (communality), and the remaining 18% would be error. The reader might also find it interesting to make an interpretation of another item, such as item 5.

Below we show an oblique solution (factor pattern) obtained with the same data. The estimated correlation between factors was 0.40. Compared to the orthogonal solution it can be seen that it is clearer and simpler (the smaller weights are now closer to zero). This is what is generally observed on comparing the two types of solution.

Ítem	F_1	f_2
X_1	0.23	**0.78**
X_2	0.18	**0.82**
X_3	0.25	**0.75**
X_4	**0.78**	0.18
X_5	**0.77**	0.14
X_6	**0.78**	0.21

The corresponding diagram would now be:



Wright Diagram for two uncorrelated common factors.

In this example we shall work with item 4. Its basic equation takes the same form as in the orthogonal case:

$$X_{i4} = 0.78\, f_{i1} + 0.18\, f_{i2} + e_{i4}$$

However, its interpretation is different, since the weights (effect of the factor on the variable) and the variable-factor correlations are different measures. Thus, the weight corresponding to the first factor is 0.78. However,

the correlation between the variable and this factor, that is, the structural coefficient, should be obtained as:

$$s_{j1} = 0.78 + 0.18 \cdot 0.40 = 0.85$$

Readers who wish to practice might like to make the calculations corresponding to item 1.

Let us now move on to the contributions of the partial correlation model. The basic equation is shown in the following box for the Spearman model, and is based on the previous results discussed above.

$$r_{jk.f} = \frac{r_{jk} - r_{jf} r_{kf}}{\sqrt{1 - r_{kf}^2}\sqrt{1 - r_{jf}^2}} = \frac{\lambda_j \lambda_k - \lambda_j \lambda_k}{\sqrt{1 - \lambda_k^2}\sqrt{1 - \lambda_j^2}} = 0 \qquad (11)$$

Equation (11) indicates that, if the model is correct, the partial correlation between any pair of variables after eliminating the influence of the general factor from both is zero. The numerator of the partial correlation is the difference between the correlation observed between the two variables and the correlation reproduced from the model, and is referred to as the residual correlation. Thus, if the model is correct, the observed correlation and the reproduced correlation are equal, and the residual is zero. Conceptually, the interpretation of this result is that the only thing the variables have in common is the general factor they measure, so that on eliminating this common cause there is no longer anything that links them. The multiple case is more complex, but the essential idea is the same. If the model is correct, and the variables have only *m* factors in common, then the partial correlation between any pair of variables after eliminating the influence of these common factors should be zero. More than a result, this is an essential principle. It suggests that the most direct way to assess whether the FA model is appropriate should be based on the assessment of the residuals after fitting a model with the proposed number of factors.

In general, most of the characteristics of the model considered up to now are dictated by the principle of parsimony (Carroll, 1978). According to the parsimony principle, the equations of the model should be linear, and thus the simplest possible. This principle also recommends making a clear distinction between common variance (communality) and error variance. Finally, this same principle suggests that the number of common factors should be considerably less than that of variables. There would be nothing to be gained, indeed, by interpreting a solution with as many factors as variables. The determination of the correct number of factors (small enough to be clearly interpretable and large enough to account for the relations between variables) is perhaps the most important decision in FA (Thurstone, 1947).

EXPLORATORY FACTOR ANALYSIS AND CONFIRMATORY FACTOR ANALYSIS

In the literature (e.g., Mulaik, 1972) a very sharp distinction is drawn between two types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). In our opinion, this distinction is not as clear as is presented in the texts; moreover, it raises a series of problems. First, in the EFA-CFA distinction, two concepts are mixed: (a) the purpose of the analysis and (b) the model that is tested. And second, in both (a) and (b), EFA and CFA are not two qualitatively distinct categories, but rather the two poles of a continuum.

As traditionally understood, in a purely exploratory analysis the researcher would analyze a set of data without having any previous hypothesis about its structure, and would leave it to the results of the analysis to provide information about it. On the other hand, in a CFA, the researcher would have drawn up a series of well-specified hypotheses that will be tested by assessing the fit of the model to the data. These hypotheses would be of three types: (a) number of factors, (b) pattern of relations between variables and factors, and (c) relations between the factors.

In initial attempts to assess a new phenomenon, a purely exploratory position such as that described above would be defensible. However, it does not seem to be so appropriate in the case of analyzing a test that we have developed or adapted ourselves. In this case, it is reasonable to suppose that we shall have a series of previous hypotheses about the number of dimensions the test is intended to measure, which items are to measure each dimension, and whether those dimensions are independent or not according to the theory. In most cases, however, these hypotheses are not sufficiently strong to allow a full CFA model to be specified. Thus, with regard to the purpose, it is useful to consider that the majority of psychometric applications of FA are to be found at some intermediate point.

As regards the type of model tested, the distinctions refer here to the degree of restriction in the solution proposed. In an EFA the restrictions imposed are the minimum ones required for obtaining an initial solution – a solution that can subsequently be transformed. In a CFA the restrictions

are much stronger, and only allow the testing of a single solution, which cannot be transformed later on. The purpose of the analysis and the type of model tested are not independent concepts. The more previous information we have and the stronger the hypotheses, the more specific will be the solution tested and the greater the number of restrictions imposed on that solution. Nevertheless, even accepting this clear relationship, the distinction between restricted FA and unrestricted FA seems to us more appropriate than that of EFA and CFA to refer to the type of model tested.

In a CFA, as normally used, the restrictions concerning the number of common factors and the relations between them are similar to those involved in an EFA. In general, the correlations between factors are freely estimated. The main differences concern the restrictions imposed on the factorial pattern. The solution almost always proposed is one called "independent clusters" (McDonald, 1985), which follows the simple structure principle. In this solution, each variable has a non-null weight in a single common factor, the weight in the remaining factors being zero. A solution of this type is presented below:

Ítem	f_1	f_2
X_1	0.0	*
X_2	0.0	*
X_3	0.0	*
X_4	*	0.0
X_5	*	0.0
X_6	*	0.0

where the asterisk indicates that the corresponding weight is estimated as a free parameter. In this hypothetical solution, the first three items would be pure measures of the second factor and would have null weights in the first. On the other hand, the last three items would be pure measures of the first factor. Solutions of this type are theoretically ideal. They have the maximum possible structural simplicity, and allow the content of each factor to be interpreted without ambiguities.

Let us look now, in contrast, at an exploratory solution obtained from the analysis of these 6 items. It was obtained from an EFA in which two common factors were proposed. The initial arbitrary solution was transformed into an oblique solution, since, in theory the two factors were considered to be related. It is the pattern we presented previously

Ítem	f_1	f_2
X_1	0.23	**0.78**
X_2	0.18	**0.82**
X_3	0.25	**0.75**
X_4	**0.78**	0.18
X_5	**0.77**	0.14
X_6	**0.78**	0.21

It is indeed quite clear, and the fit to the two-factor model was good. It seems fairly evident that the first three items mainly measure f_2 and the last three f_1. However, are these items "clean" enough to fit well with the previous hypothetical solution?

If we assess the fit of the data to the hypothetical solution presented above, that is, we fit a conventional CFA to these items, what we are proposing is that each one of them is a pure measure of a single factor, and therefore, that the lower weights which appear in the EFA solution are due only to sampling error and are, therefore, compatible with values of exactly zero in the population.

The problem with this approach is that, in the real world, the majority of items (and of tests) are not factorially pure measures. With some effort and after a process of selection it is possible to obtain some items that are (almost) pure measures. Such items are called "markers" or "indicators" in the language of FA. Nevertheless, expecting all the items in a test to be markers seems to us a somewhat unrealistic hypothesis.

If we accept the idea that in the majority of FAs many of the items are factorially complex, we will have to conclude that the structural hypothesis most common in a CFA is false, and that, therefore, the model will not fit well. More specifically, if the lower weights (those around or under 0.20) are low but not null, every time we fix one of them at zero we commit an error of specification of the model. If the model has few items, such as in the example, we can perhaps still attain an acceptable fit. However, in larger models, the accumulation of errors will necessarily lead to unacceptable goodness-of-fit values. This reasoning explains two results that give cause for considerable concern in the applied field (e.g., McCrae, et al. 1996). The first is that factorial structures obtained by means of EFA that are clear, interpretable and replicable across different studies show inadmissible goodness-of-fit values when assessed by means of CFA. The second is that it is easier to obtain a poor fit when analyzing questionnaires

of realistic size than when analyzing very small groups of items. The first result may lead to lack of confidence in FA on the part of the researcher; the second may lead to ill-advised practices and to the unnecessary elimination of items.

Our position can be summarized as follows. In the analysis of items and tests, we believe that FA should be guided by previous theory. This theory will allow the proposal of hypotheses about the number of factors, the (approximate) pattern we expect to find, and whether the factors are related or not. Nevertheless, in general, prior knowledge will not suffice to specify a confirmatory model. What we propose is to use a non-restrictive (exploratory) model, but with a confirmatory purpose as far as possible. That is, to estimate a solution in which the number of factors (or at least a range of values) is specified, and also whether these factors are independent or not. Moreover, we should have a more or less clear idea of the nature of the transformed pattern that will be obtained. Naturally, if the study is advanced enough to be able to specify a restrictive solution, or all the items are exceptionally simple, then CFA is the model to use.

DESIGN OF A STUDY BASED ON FACTOR ANALYSIS

As in any statistical analysis, for the results obtained through FA to be valid, interpretable and generalizable, some basic conditions must be fulfilled in the research design. To clarify the importance of this point in our case, it is useful to consider FA as an analysis at two levels. At the first level the correlations between a series of measures are calculated. At the second level the structure of these correlations is analyzed. If the results are already wrong at the first level, they can never be correct at the second level. For reasons of clarity, we shall discuss separately the two basic aspects in the design: sample and variables.

SAMPLE

In any factorial study, and especially in those involving the development or adaptation of a test, we should have a relatively clear idea of the population of interest. Thus, FA should be based on a representative sample of this population. It is customary, however, to use samples of convenience (usually university students). Apart from the lack of representativeness of such samples, the most substantial statistical problem here is that of attenuation due to restriction of range. If the sample is highly homogeneous as far as the variables to be analyzed are concerned (i.e., if the scores on the items/tests have little variability), then the correlations obtained at the first level of FA will be attenuated. The correlation matrix will then have much more "noise" than "signal", and it will be difficult to obtain a clear solution at the second level.

Perhaps the most widely discussed problem in FA in relation to the sample is the stability of the solution (How large a sample is needed for a solution to be stable and generalizable?). This is a complex problem. The stability of a factorial solution depends jointly on three factors: (a) the sample size, (b) the degree of determination of the factors, and (c) the communality of the variables. Thus, if the factors are well determined and the variables have little measurement error, stable solutions can be achieved with relatively small samples. In this regard, we should stress that traditional "recipes" of the type "10 times more participants than variables" lack a solid basis.

The measures normally used in psychology – tests, and above all, items – contain a great deal of intrinsic measurement error. It must be accepted, then, that the communalities will be generally low, and that therefore, we should act primarily on points (a) and (b). With regard to point (b), which is discussed in detail below, the idea of determination of a factor refers to the number of variables with high weights in that factor – i.e., the extent to which the factor is well defined and clearly measured by a substantial number of indicators. As regards point (a) it is appropriate once more to think in terms of a "dual level". The results of the second-level analysis can only be stable if the correlations on which they are based are themselves stable (and it should be stressed that correlations have high sampling fluctuations). It is therefore reasonable to consider a sample of 200 observations as a minimum, even in ideal circumstances (high communalities and well determined factors).

VARIABLES

FA is a model for continuous and unlimited variables. But neither item scores nor test scores have these qualities. Therefore, in the majority of psychological applications, FA should be seen as an approximate model whose advantage lies in its simplicity. It is important, then, to first of all discuss in which conditions the approach will be sufficiently sound for what is required in practice.

FA generally works well in the analysis of scores on tests and subtests. As for items, this approach also tends to be acceptable when graded response scales (Likert) with 5 or more categories are used. Finally, binary items and items

with 3 options and a central category can cause more problems. In principle we would recommend using the graded response format wherever possible.

Whatever the type of response, whether or not FA works well depends above all on the distribution of the scores. Symmetrical distributions do not tend to give problems. On the other hand, the most significant problems occur when (a) the distributions are markedly asymmetrical, and (b) the asymmetries are in both directions. An example of this situation would be the analysis of a test containing very easy items or very difficult items. Asymmetries of opposite signs give rise to non-linear relations, and therefore, to unsuitability of the linear FA model (Ferrando, 2002). In relation to what was set out above, the magnitude of the problem depends on the type of variable to be analyzed. With test scores it is very difficult for non-linear relations to occur. With Likert items it is a problem to take into account. Finally, it is a very common problem in binary items, and known as the 'difficulty factors' problem (McDonald & Alhawat, 1974).

Results obtained in simulation, together with actual experience, lead us to the following recommendations. In the case of tests and subtests, FA is nearly always appropriate. In the case of graded response items, FA is expected to work well if the asymmetry coefficients are all in the interval between -1 and +1. Finally, even binary items can be fitted well by the linear model if the difficulty indices are in the range 0.4 to 0.6. When the variables have more extreme distributions, it is generally necessary to resort to non-linear approaches, which we are unable to deal with here.

Apart from the scale and the distribution, there are other factors to take into account as far as variables are concerned, especially where individual items are involved. As we have said, the reliability of the items is intrinsically low. However, the analysis of items with excessively low reliabilities should be avoided, since such items will only add noise to the factorial solution. A conventional pilot study assessing the discrimination indices (item-total correlations) or the test-retest correlations item by item is highly recommended. This allows us to eliminate those items that contribute only noise and to begin the FA from a cleaner input.

It is relatively common for typical performance questionnaires (personality, motivation and attitudes) to include redundant items – that is, those which are essentially the same question put slightly differently. Such items are used for assessing the consistency of the participants or (in overlapping form) to increase the internal consistency of the test. The presence of redundant items always gives rise to problems in FA. In fact, the errors between two redundant items cannot be independent, since, even after eliminating the common factors, the responses continue to be related due to the similarity of the content. The consequence is the need to extract additional factors defined principally by pairs or "triplets" of redundant items. These factors may be difficult to identify, particularly in rotated solutions. A prior content analysis can eliminate redundancy and avoid such problems from the outset.

Finally, we shall consider the degree of determination of the factors. Where possible, a good recommendation is to use markers or indicators. As we have already said, markers are, theoretically, pure measures of a factor. In more applied terms, Cattell (1988) defines them as variables which, in previous studies, have shown themselves to be good measures of the factors under assessment. Their use has two main functions: (a) it allows us to identify the factors by increasing their degree of determinacy, and (b) it permits us to relate the results of the study to those of previous work. Cattell (1988) recommends using a minimum of two markers per factor.

As regards the relationship between the number of items and the number of factors, as we know, the more items there are that accurately measure a factor, the more well-determined that factor will be, and the more stable the solution. Although divergent recommendations can be found (Cattell, 1988), our opinion is that the best designs in FA are those in which few factors are posited, markers are employed and a substantial number of items is proposed for measuring each factor. Whether or not markers are used, in order to identify a factor clearly it is necessary to have a minimum of 4 variables with substantial weights on that factor.

STAGES OF A FACTOR ANALYSIS

Preliminary analyses: adequacy of the data

In accordance with the dual-level approach, it would appear logical, before undertaking an FA, to use indicators for assessing whether the correlations obtained at the first level are appropriate for factorial analysis at the second level. These indicators are usually referred to as "sampling adequacy measures", and their use is highly important as a preliminary stage of FA: it will indicate whether the FA is or is not the appropriate model for the data. However, this is the stage that is most neglected in applied research.

To begin with, it is advantageous to inspect the descriptive statistics of the variables, in accordance with what was discussed in the previous section. Next, Bartlett's (1950) sphericity test should be carried out. This constitutes a test of the null hypothesis that the population correlation matrix is identity – that is, that the variables are uncorrelated in the population. If this hypothesis cannot be rejected, we must accept that the correlation matrix contains only "noise". It is important to take into account that, all the same, if this matrix is analyzed, factors will be obtained. However, such factors will be totally spurious. In this regard, it is useful to consider the Bartlett test as a security measure and a necessary condition. In most FAs the null hypothesis will be rejected, and thus it will be admitted that there is some relationship between the variables. However, this may not be sufficient. As we have seen, the FA model assumes, moreover, that the relationship is substantial. If the relationship is so diffuse that we need practically as many factors as variables to explain it, then it is not worth carrying out the analysis.

Assuming that the necessary condition is fulfilled, in third place the degree of joint relation between the variables would be assessed. The most widely used measure is the KMO (Kaiser, 1970), which assesses the degree to which the scores in each of the variables can be predicted from the others. The range of values of the KMO is 0 to 1, and the higher the value, the more substantially related between one another are the variables. As a value of reference, Kaiser (1970) suggests that the correlation matrix will be appropriate for factorization if the KMO is 0.80 or higher.

Estimation of the model

As mentioned earlier, this is the crucial stage of FA. It involves the estimation of an initial solution and, above all, determination of the dimensionality of the data, that is, the most appropriate number of factors. The estimation stage should be guided by the principle of parsimony. The aim is to determine the simplest solution (i.e., with the lowest number of factors) compatible with residuals sufficiently close to zero.

The estimation procedure implemented by default in statistical programs is usually principal components analysis (PCA). PCA, however, is not a procedure for estimating the factorial model. It is a method for reducing the number of variables. In essence, FA is a model which is based on the principle that the variables have measurement error, which distinguishes clearly between common variance (communality) and error variance, and which sets out to reproduce only the common variance – that which is involved in the correlations between variables. PCA, on the other hand, does not make such a distinction, considers only total variance and sets out to reproduce the total variance.

Advocates of PCA argue that it is simpler, better defined and produces virtually the same results as FA (e.g., Velicer, 1990). However, the last claim is only half true. Theoretically, and from the point of view of FA, PCA could be considered as the extreme case of the factorial model in which all the variables for analysis are error-free (i.e., common variance and total variance coincide). In practice, PCA and FA lead to similar results when: (a) the number of variables to be analyzed is large (say, over 30) and (b) the variables have little error, and therefore, high communality (Mulaik, 1972). A basic principle in psychometrics, however, is that test scores have measurement error (and item scores have much more). It does not seem very reasonable, then, to use a technique not based on this principle.

The problem with using PCA when the correct model is FA can be illustrated by a small simulation. A correlation matrix was generated on the basis of the following "true" factorial solution

0.50
0.50
0.50
0.50
0.50
0.50
0.50
0.50
0.50
0.50

Next, the correlation matrix was analyzed by means of a proper factorial method and by means of PCA. The direct factorial pattern (two factors being specified) was:

0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00
0.50	0.00

which reproduces exactly the true solution. In contrast, the PCA solution was:

0,57	0	-0,82	0	0	0	0	0	0	0
0,57	0,03	0,09	-0,1	-0,8	0,22	0,03	-0,02	-0,01	-0,02
0,57	-0,04	0,09	-0,6	0,01	-0,47	-0,12	0	-0,1	-0,04
0,57	0,4	0,09	0,24	0,01	-0,34	0,44	0	0	0,34
0,57	0,15	0,09	0,18	0,12	-0,02	0,13	-0,15	-0,06	-0,74
0,57	0,06	0,09	0,18	0,12	0,13	-0,33	0,33	-0,59	0,1
0,57	0,09	0,09	0,24	0,01	-0,14	-0,55	0	0,51	0,05
0,57	0,03	0,09	-0,3	0,23	0,41	0,24	0,45	0,32	0
0,57	-0,73	0,09	0,24	0,01	-0,14	0,19	0,01	0,01	0,05
0,57	0	0,09	-0,1	0,23	0,35	-0,02	-0,63	-0,06	0,24

which shows the two problems typical of PCA: upwardly biased estimations of the weights in the content factor and overestimation of the dimensionality. The first component is a biased estimator of the only 'real' factor (whose 'true' weights are all 0.50). On the other hand, in the successive components some of the variables have weights of over 0.20-0.30. In this regard it is important to note that, in practice, it tends to be recommended to interpret only the weights that are above these minimum values (Catell, 1988, McDonald, 1985). McDonald (1985) proposes a more restrictive heuristic criterion according to which only those factors with at least three variables with weights over 0.30 would be interpreted. Even following this more restrictive criterion, the results of the PCA solution would lead to the interpretation of 4 components as if they were true factors, while, in fact they reflect only error. If, moreover, this solution had been rotated subsequently, it would possibly have been totally erroneous.

Various methods are recommended for estimating the FA model. For reasons of space, we shall discuss here only the two most widely used: Ordinary Least Squares (OLS) and Maximum Likelihood (ML). However, there are other very interesting methods whose study we recommend to the interested reader. Particularly worth reviewing is minimum rank FA (Shapiro & ten Berge, 2002).

FA by OLS is not, strictly speaking, a method of estimation, but rather a set of methods based on a common general criterion. For the specified number of factors, OLS estimators are those that minimize the sum of the squared differences between the observed correlations and those reproduced from the model. Conceptually, then, OLS methods determine the solution that makes the residuals as close to 0 as possible. This is, as we know, the basic idea of fit in FA. Although the criterion is very clear and direct, OLS methods are, in principle, purely descriptive. As we shall see, however, this is not necessarily a limitation.

The principal methods based on the OLS criterion are: (a) principal axes FA, (b) Harman's MINRES (Harman & Jones, 1966), (c) Jöreskog's (1977) ULS, and (d) Comrey's (1962) Minimal Residual. For the same number of factors, the solutions obtained with any of them are virtually identical. Nevertheless, we would especially recommend the use of MINRES or ULS for two reasons: (a) they do not require initial estimation of the communalities and (b) they are more efficient in terms of computation.

In contrast to the OLS methods, the ML method (Lawley & Maxwell, 1971) is statistical (inferential). Its main advantage is that it permits the rigorous examination of the model's fit to the data through an index referring to the chi-squared (χ^2) distribution. This advantage, however, should be put into context. First of all, the inference in ML FA is based on the assumption that the variables analyzed are continuous, metric and distributed according to the normal multivariate law. In the case of items and tests, this assumption is never met. Secondly, it is assumed that the model proposed in *m* factors fits perfectly in the population, so that all of the error is sampling error (this is the null hypothesis of the goodness-of-fit test). However, as we saw previously, models are proposed only as reasonable approximations, and it is accepted that part of the error will be error of approximation. Thus, the method assesses a null hypothesis that we know from the outset to be false, and which will always be rejected as soon as there is sufficient power. To make matters worse, the power is generally very high, since FA normally involves working with large samples. In sum, even with 'reasonable' distributions, the use of ML FA based on the goodness-of-fit test will nearly always lead to the need to estimate more factors than are substantively interpretable. This phenomenon is called "over-factorization".

In spite of the problems referred to above, however, there are reasons to recommend the use of ML FA. First of all, and although it is not well known, the ML solution can also be obtained without making inferential assumptions. It is that which minimizes the partial correlations between the variables after eliminating from them the influence of the factors (Mulaik, 1972). Essentially, it is the same basic criterion as that of FA by OLS. OLS methods minimize the residual correlations. ML minimizes the partial

correlations (the residual correlation is the numerator of the partial correlation). For this reason, in practice, OLS and ML solutions tend to be very similar. Secondly, although the χ^2 goodness-of-fit statistic evaluates a false hypothesis, there are goodness of fit indicators derived from this statistic that assess the error of approximation and the model's degree of fit. As we shall see in the following section, these indicators are highly recommended.

In a situation in which (a) the variables have acceptable distributions, (b) the solution is well determined, and (c) the model proposed is reasonably correct, the OLS and ML solutions will be practically identical. In this case, the use of ML has the advantage of permitting us to obtain additional indicators that are very useful in the assessment of fit. In the case of extreme distributions, weak or unclear solutions and high error of approximation, the ML option will give problems. The convergence of the method is quite delicate, and may give rise to unacceptable estimations. Moreover, the additional indicators will not be reliable. In these cases the OLS methods are clearly superior. According to simulation studies, they are highly robust methods (they nearly always converge), and since they do not make distinctions between error sources, they tend to approximate the correct solution more effectively than the ML method (MacCallum & Tucker 1991).

Assessment of fit

In order to decide whether a model with *m* factors is appropriate, it is necessary to assess the model's degree of fit to the data. There are a variety of criteria and procedures for performing this assessment. In our view, some are considerably better than others.

Let us begin with the criteria and procedures we do not recommend. Possibly the most widely used criterion in Psychology, and that which commercial programs tend to apply by default, is Kaiser's criterion: the relevant number of factors is the number of eigenvalues greater than 1 in the original correlation matrix. This criterion presents several problems, the first of these being the lack of a clear justification. It has several (which we shall not see here), but none of them convincing. The second problem is that it is based on the logic of PCA, not on that of FA. Indeed, the eigenvalues of the matrix without reduction (with ones in the principal diagonal) are equivalent to the proportions of total variance explained by the corresponding principal components. However, as we have seen, the variance of real interest in FA is the common variance, not the total variance. Thirdly, the number of factors determined by means of this rule is related to the number of variables analyzed. In more detail, if *n* variables are analyzed, Kaiser's criterion will indicate a number of factors ranging between n/5 and n/3. Thus, with 30 variables, the criterion will indicate between 6 and 10 factors. However, if we have designed a scale of 30 items to measure a single dimension, the expected number of factors is 1, not between 6 and 10.

The scree test (Cattell, 1988) is a widely used graphical procedure. On a bivariate graph, points are plotted whose coordinates are the eigenvalues of the original correlation matrix (i.e., the proportions of total variance explained) on the y axis and the number of components on the x axis. In a typical solution, the graph linking the points is a decreasing function, similar in form to a scree slope. From a certain point onwards the function becomes practically horizontal, and it is this point which, according to Cattell, indicates the most appropriate number of factors. The logic is that, from this number onwards, the successive factors are trivial, and only explain residual variance. Although the logic is more convincing than that of Kaiser's rule, there are, in our view, two problems with the procedure. First, the decision is based on visual inspection, and therefore has a strong element of subjectivity. Secondly, it is based on the logic of PCA, and fails to distinguish between common variance and error variance. Even so, if instead of the eigenvalues of the matrix without reduction (proportions of total variance), those of the reduced matrix (i.e., the communalities) were represented, then the test would be useful as an auxiliary procedure.

Two criteria that are highly fashionable just now are Velicer's MAP (1976) and parallel analysis (PA; Horn, 1965). In our opinion they are useful as auxiliary criteria, but are affected by the same basic problem as the previous criteria based on the logic of PCA: they fail to distinguish between common variance and error variance.

In the MAP criterion, a PCA is carried out in sequential form, and at each stage one calculates the root mean square of the partial correlations resulting from elimination of the corresponding component and the previous ones. Although the residual correlations always decrease as more components are estimated, the partial correlations do not do so. In fact, the function relating the root mean square of the partials with the number of components tends to take the form of a U. The minimum

of the function would indicate the number of components to be retained.

Parallel analysis (PA) can be understood as a combination of Kaiser's criterion and the scree test. In its most basic form, it consists in generating a random correlation matrix based on data from the same dimension as the empirical data (participants and variables): theoretically, such a matrix should have all the eigenvalues close to 1. The method consists in comparing the eigenvalues of the empirical matrix with those of the randomly-generated matrix. In graphical terms, the comparison would be like a double scree test involving the simultaneous representation of the curve corresponding to the empirical data and that corresponding to the random data. The former, as we know, would be expected to show a strong fall followed by stabilization. The second should show a much flatter tendency (in a very large matrix it would be a horizontal straight line cutting the y axis at 1). The point of intersection between the two curves would indicate the number of factors to be retained. Considered in this way, many of the same criticisms can be levelled at PA as in the case of the scree test. However, PA does have the advantage that the criterion for determining the number of factors is much more objective.

Let us now move on to a discussion of the criteria and procedures we consider most worthy of recommendation. To begin with, we shall consider those of a general nature that can be applied whatever the method of estimation. Subsequently we shall discuss those specifically related to ML FA.

As we already well know, if the number of factors proposed is appropriate, then the residual correlations between the variables after eliminating the influence of the factors should all be practically zero. In accordance with this principle, the clearest criteria for assessing the fit of a model in *m* factors will be those most directly related to the assessment of the residual correlations.

In small problems, visual inspection of the matrix of residuals can already give a good idea of the degree of fit. However, FA usually works with a substantial number of variables, so that global inspection of the residual matrix is impractical. In such a case, the information should be condensed by means of descriptive statistics.

To begin with, it is useful to inspect the distribution of frequencies of the residuals. If the number of factors proposed is adequate, this distribution will be symmetrical, approximately normal and centred around a mean of zero. Asymmetrical, off-centre or very heavy-tailed distributions indicate that there remains some systematic covariation to explain, so that it is necessary to estimate more factors.

The root mean square residual (RMSR) is a descriptive measure that indicates the average magnitude of the residual correlations. If their mean is zero, then the RMSR coincides with the standard deviation of the residuals. Harman (1976) proposes a value of reference of 0.05 or less for considering that the model's fit is acceptable. This criterion is purely empirical, but generally functions well in practice. More well-founded is the criterion initially proposed by Kelley (1935). The standard error of a correlation coefficient of zero in the population is approximately $1/\sqrt{N}$, where N is the sample size. This would be, therefore, the value of reference. Thus, in a sample of 300 participants, $1/\sqrt{N}=0.058$. If the RMSR is around this value, or is lower, we can interpret that the residual values observed do not differ significantly from zero, and therefore, that there are no longer any systematic relations to be explained.

The gamma index or GFI initially proposed by Tanaka and Huba (1985) is a goodness of fit measure normed between 0 and 1 that can be used with the majority of estimation procedures. It can be interpreted as a multivariate determination coefficient that indicates the proportion of covariation between the variables explained by the model proposed. According to the current criteria (see the article by Ruíz, Pardo and San Martín in this issue), for a fit to be considered as good, the GFI should be above 0.95.

Finally, we consider two indicators that are used in the case of ML estimation. They are, therefore, inferential, but here we recommend their use in accordance with descriptive logic. The first of these is the TLI-NNFI coefficient, initially proposed by Tucker and Lewis (1973), precisely for the FA model. It is a relative index, and measures the improvement in fit produced by the model proposed with respect to the null model in 0 factors, in relation to the improvement expected by a model with good fit. Its values range between 0 and 1 (even though it is not strictly normed), and Tucker and Lewis recommend interpreting it as a reliability coefficient. In this sense, and although the current criteria are more rigorous (see Ruíz, Pardo and San Martín in this issue), our opinion is that values above 0.85-0.90 would start to be considered acceptable.

The RMSEA index (Steiger & Lind, 1980), very much in vogue at the moment, estimates the error of

approximation of the model proposed. More specifically, it estimates the discrepancy there would be between the populational correlation matrix and the matrix reproduced by the model proposed, also in the population. Conceptually, RMSEA is based on the approach, discussed above, that models are only approximations, and estimates the extent to which the model tested is a reasonable approximation. RMSEA is an index relative to the degrees of freedom (complexity) of the model, and may therefore penalize less parsimonious models. As a reference, values below 0.05 could be considered indicators of good fit, whilst values between 0.05 and 0.08 would indicate acceptable fit.

We conclude this section with three observations. First, we do not recommend making a decision as important as that of the appropriate number of factors based on a single indicator or criterion. It is appropriate to use multiple indicators that provide us with more information, and therefore, with more elements for making a judgement. Second, in real situations the process leading to the determination of the number of factors is not as linear as described here for didactic reasons. In fact, having obtained the transformed solution we may find that one or more factors are very weak, are poorly identified, or reflect trivial content (e.g., they do not attain the minimum of 3 variables with weights over 0.3, mentioned previously). This result could lead to the reconsideration of the number of factors and a further inspection of the solution. And finally, in many cases the previous theoretical information does not indicate a clear number of factors, but is actually much less specific. It may indicate a plausible range of factors or perhaps various plausible alternative solutions. It is advantageous in this case to examine the differences between the values of the fit indicators corresponding to the different solutions assessed.

Obtaining the transformed solution (rotation)

In unrestricted factor analysis, the standard initial solution obtained by means of applying OLS or ML methods is referred to as "canonical form". For the specified number of factors, this solution has the property that the successive factors explain the maximum possible amount of the common variance. In some cases the canonical solution is directly interpretable, and requires no subsequent transformation. The clearest case is that in which a set of items is analyzed under the hypothesis that they measure a single dimension. Given that the first factor explains the maximum possible amount of common variance, if the set is essentially unidimensional, then the factors that follow the first one must be residual. An extreme example is the direct two-factor solution we have used above to compare FA with PCA. It is a canonical solution in which the first factor already explains all the common variance, leaving nothing for the second factor to explain. This group of variables would therefore be perfectly unidimensional.

When we expect to find a multiple-factor solution, however, the initial canonical solution is arbitrary. It must then be transformed or rotated until a solution is obtained that can be interpreted in accordance with the theory on which the analysis is based.

The principal decision the researcher has to make at this stage is whether to use oblique or orthogonal rotation. This is a controversial issue. Authors who advocate orthogonal solutions consider them simpler and easier to interpret. Moreover, they believe they are also more stable in replication studies. The statistical basis of this argument is that if the factors are independent in the population, they will not be exactly so in the samples, and therefore, if oblique solutions are used, the correlations between factors will reflect only sampling error. On the other hand, authors who defend oblique solutions consider that the majority of psychological constructs are related, and that to specify uncorrelated factors is to artificially impose an incorrect solution only because it is simpler (e.g., Mulaik, 1972). In sum, it is theory that must guide this decision.

The authors of the present work essentially espouse the second view. However, we also believe it necessary to take into account (as always in FA) the criterion of parsimony. If the theory does not permit strong hypotheses, it would seem reasonable to begin with an oblique solution. If the correlations estimated between factors are substantial, this is the solution to maintain. However, if the correlations between factors are consistently low (say, under 0.30 or 0.20), then a second analysis may be made, specifying an orthogonal solution. Should the two solutions be similar, it would be preferable to provisionally accept the orthogonal solution.

Once the general type of rotation has been decided, the importance of the decision about the specific method to employ will depend on the solidity of the design. If the

variables are 'clean', markers are used and the factors are well determined, then the different methods should lead essentially to the same solution. It is not a bad strategy to try out different methods and assess whether they converge approximately in a common solution. In complex, noisy and poorly determined solutions, the use of different rotation methods can lead to highly disparate solutions. This is not a problem of methods, but rather one of design.

Analytical methods of orthogonal rotation are generally 'quartic', in the sense that they are based on the quartic functions of the factorial weights (conceptually, variances of the squares of the weights). There are two general methods of this type (see, for example, Mulaik, 1972): Quartimax and Varimax. Given the initial unrotated pattern (variables x factors), Quartimax transformation is that which maximizes the variance of the squares of the weights by rows. In contrast, Varimax maximizes the variance by columns. The Quartimax solution is therefore compatible with a column of the pattern in which the majority of the weights are high, and tends to give solutions with a general factor. On the other hand, Varimax tends to give multiple solutions in which there is no dominant factor. There is, in fact, a third method, Equamax, which combines the two criteria and therefore leads to intermediate solutions. All three methods function well, and the choice of one or another should perhaps be guided by what is expected from the theory (whether or not there is a general factor). Our positive consideration of the methods (especially Varimax) is not incompatible with the critique we make at the beginning of the article, which refers to the indiscriminate use of Varimax by default even when the theory clearly indicates that the factors are related or when it is expected to find a general factor.

Experience with the above methods based on simulation studies suggests that they may lead to erroneous or unstable results (especially Varimax) when a high proportion of the variables to be analyzed are factorially complex. As we stated above, this is not a problem of the rotation method, but rather one of poor design. In order to minimize the problem, weighted versions of the three methods (Quartimax, Varimax and Equamax) have been proposed in which greater weight is assigned to the items initially assessed as the simplest.

Traditional analytical methods of oblique rotation are an extension of orthogonal quartic methods, with the additional problem of determining the degree of relation (obliqueness) between the factors. In practice, by far the most widely used method is Oblimin (see, for example, Mulaik, 1972) whose criterion can be considered as a combination of the Quartimax and Varimax criteria extended to the oblique case. The Oblimin criterion includes a parameter (delta, which can take values between 0 and 1) that controls the maximization of the simplicity by rows or by columns. Indirectly, the delta parameter also controls the degree of obliqueness permitted in the solution. Factorial programs use by default the value delta=0, following Jennrich's (1979) recommendation. This value tends to lead to good convergence and to a solution that is simple and interpretable. On the other hand, in order to achieve this simplicity, the resulting factors tend to be closely related to one another. Browne (personal communication) recommends using delta=0.5.

An interesting alternative to Oblimin would be certain analytical methods that incorporate the idea of rotation on a target matrix that is constructed analytically. Essentially, the idea is as follows. First of all, from an orthogonal solution, a hypothesis or target matrix is constructed. This matrix is a modification of the orthogonal solution which approaches as closely as possible the simple structure criterion; thus, and principally, the very low weights in the orthogonal solution are hypothesized as zeros in the target matrix. Secondly, the oblique transformed solution is determined which most closely approximates to the target matrix. The initial method that incorporated these ideas is Promax (Hendrickson & White, 1964); Lorenzo-Seva (1999) recently proposed a much more refined method called Promin. With respect to the previous methods, Promin incorporates improvements at all stages: initial orthogonal solution, determination of the target matrix and procedures and criteria for obtaining the best approximation. It is, therefore, the most highly recommended method within this family.

SOFTWARE: THE FACTOR PROGRAM[1]

Whilst it is true that EFA is a classic data analysis technique, statistical research on analysis itself is constantly progressing. Thus, recent years have seen the publication in specialist journals of a range of advances

[1] The author of this section is Urbano Lorenzo-Seva

related to the methods employed in EFA. However, the authors of the most popular data analysis programs (it is not necessary to mention names) do not appear to be interested in implementing these new advances. For this reason, university researchers have taken it upon themselves to develop specific programs that incorporate both classic methods and new contributions. An example of such programs is Factor (Lorenzo-Seva & Ferrando, 2006). This is a program that is easy to use (based on the typical Windows menus), and whose purpose is to fit the EFA model.

Factor implements classic procedures and indices, as well as some of the most recent methodological contributions – including not only all of those discussed in this article but also others of undoubted interest that should be the object of future study for readers interested in FA. Thus, for example, Factor makes it possible to work with polychoric correlations when it is suspected that the linear model might be inadequate. Good examples of the methodology recently proposed that have been implemented in Factor are: (1) Minimum Rank Factor Analysis, which is the only factor-extraction procedure that permits assessment of the proportion of variance explained by each factor; and (2) the Simplimax rotation method, which has shown itself to be the most efficient rotation method of all those proposed up to now. Many of these methods are not available in any commercial program.

Finally, it should be stressed that Factor is a program distributed free of charge (i.e., free software). It can be obtained in Internet at: http://psico.fcep.urv.es/utilitats/factor/. The same page offers a brief user's manual, as well as a demonstration version. To date, since being made available in 2006, it has been used by a wide range of researchers in 29 international articles published in ISL journals.

ILLUSTRATIVE EXAMPLE

Below, we propose a practical exercise for the reader to apply the material presented. The data can be found at the following address: http://psico.fcep.urv.cat/ejemplopapeles. The 14-item test measures two types of anxiety, and the general aim is that, by using the Factor program, the reader carries out the factor analysis of the test to determine its structure and properties. The solution corresponding to the example can also be found there, though we recommend that the reader works on the example before reading the solution.

RECOMMENDED READING

- **(Manuals of a general nature).** Those by Harman (1976) and Mulaik (1972), both in the list of references, deal in quite in-depth fashion with the principal aspects of FA, including those we have not been able to cover here. Their technical level is considerably higher than that of the present article.
  The manual by McDonald (1985, in the list of references) is much more personal than the two previous ones. It is highly critical of the traditional FA approach, and reflects the author's strong positions. Highly recommended.
  Finally, the following text by one of the present authors: Ferrando, P.J. (1993) *Introducción al análisis factorial* (Introduction to factor analysis). Barcelona: PPU.
  This can be used as a more technical extension of the basic aspects dealt with in the first sections.
- **The EFA vs CFA problem in item analysis is discussed extensively in**:
  Ferrando, P.J. and Lorenzo-Seva, U. (2000).
  "Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions", *Psicológica*, 21, 301-323.
- **The following thesis offers a comprehensive review of the principal criteria and methods of rotation:**
  Lorenzo-Seva, U. (1997). *Elaboración y evaluación de métodos gráficos en análisis factorial* (Development and assessment of graphical methods in factor analysis). Universidad de Barcelona.
- **Finally, the step-by-step development of an FA by means of a computer program is described in:**
  Ferrando, P.J. and Lorenzo, U. (1998)
  Análisis factorial (Factor analysis).
  In: Renom, J. (Coord.). *Tratamiento informatizado de datos* (Computerized data processing). Chap. 5, pp. 101-126. Barcelona. Masson.

REFERENCES

Bartlett, M.S. (1950). Tests of significance in factor analysis. British *Journal of Mathematical and Statistical Psychology, 3*, 77-85.

Carroll, J.B. (1978). How shall we study individual differences in cognitive abilities? Methodological and theoretical perspectives. *Intelligence, 2*, 87-115.

Cattell, R.B. (1988). The meaning and strategic use of factor analysis. In J.R. Nesselroade and R.B. Cattell (eds.) *Handbook of multivariate experimental psychology* (pp.

131-203). New York: Plenum Press.

Comrey. A.L. (1962). The minimum residual method of factor analysis. *Psychological Reports, 11*, 15-18.

Ferrando, P.J. (1993) *Introducción al análisis factorial*. Barcelona: PPU.

Ferrando, P.J. & Lorenzo-Seva, U. (1998). Análisis factorial [Factor analysis]. In: Renom, J. (Coord.) *Tratamiento informatizado de datos*, Cap. 5 [Computerized data processing, Chap. 5] (pp. 101-126). Barcelona: Masson.

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

Ferrando, P.J. (2002). Theoretical and empirical comparisons between two models for continuous item responses. *Multivariate Behavioral Research, 37*, 521-542.

Harman, H.H. & Jones, W.H. (1966). Factor analysis by minimizing residuals (minres). *Psychometrika, 31*, 351-368.

Harman, H.H. (1976). *Modern factor analysis*. Chicago: Univ. of Chicago press.

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.

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

Jennrich, R. I. (1979). Admissible values of delta in direct oblimin rotation. *Psychometrika, 44*, 173-177

Jöreskog, K.G. (1977). Factor analysis by least-squares and maximum-likelihood methods. In: K. Enslein, A. Ralston, & H.S. Wilf (Eds.): *Statistical methods for digital computers*. (pp. 125-153). New York: Wiley.

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, 26* (p. 146). Cambridge: Harvard University press.

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

Lorenzo-Seva, U. (1997). *Elaboración y Evaluación de Métodos Gráficos en Análisis Factorial* [Development and Assessment of Graphical Methods in Factor Analysis]. Doctoral dissertation. Universidad de Barcelona.

Lorenzo-Seva, U. (1999). Promin: A Method for Oblique Factor Rotation. *Multivariate Behavioral Research, 34*, 347-365.

Lorenzo-Seva, U. & Ferrando P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Reserch Methods, 38*, 88-91.

MacCallum, R.C. & Tucker, L.R. (1991). Representing sources of error in the common factor model: implications for theory and practice. *Psychological Bulletin, 109*, 502-511.

McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H. & Paunonen, S.V. (1996). Evaluating replicability of factors in the revised NEO personality inventory: confirmatory factor analysis versus Procustres rotation. *Journal of Personality and Social Psychology, 70*, 552-566.

McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale: LEA.

McDonald, R.P. and Ahlawat, K.S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27*, 82-99.

Mulaik, S.A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.

Shapiro, A. & ten Berge, J.M.F. (2002). Statistical inference of minimum rank factor analysis. *Psychometrika, 67*, 79-94.

Spearman, Ch. (1904). General intelligence; objectively determined and measured. *American Journal of Psychology, 115*, 201292.

Steiger, J.H. & Lind, J. (1980). *Statistically based tests for the number of common factors*. Paper presented at the Annual Meeting of the Psychometric Society. Iowa City, May 1980.

Tanaka, J.S. & Huba, G.J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology, 38*, 197-201.

Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago: University of Chicago press.

Tucker, L.R. & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.

Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika 41*, 321–337.

Velicer, W.F. & Jackson, D.N. (1990). Component analysis versus common factor analysis: Some further observations. *Multivariate Behavioral Research, 25*, 97-114.