LATENT VARIABLES AND FACTOR ANALYSIS

VOLUME II

SAGE BENCHMARKS IN
SOCIAL RESEARCH METHODS

SAGE BENCHMARKS IN SOCIAL RESEARCH METHODS

LATENT VARIABLES AND FACTOR ANALYSIS

VOLUME II

Exploratory and Confirmatory Factor Analysis

Edited by

Salvatore Babones

SAGE reference

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1, Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Chris Rojek
Assistant editor: Colette Wilson
Permissions: Enid Andrews
Production controller: Bhairav Sharma
Proofreader: Asish Sahoo
Marketing manager: Teri Williams
Cover design: Wendy Scott
Typeset by Zaza Eunice, Hosur, India
Printed and bound by CPI Group (UK) Ltd,
Croydon, CR0 4YY [for Antony Rowe]

FSC www.fsc.org MIX Paper from responsible sources FSC® C013604

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the Egmont grading system. We undertake an annual audit to monitor our sustainability.

First published 2015

Library of Congress Control Number: 2014943396

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN: 978-1-4462-9460-4 (set of four volumes)

LATENT VARIABLES AND FACTOR ANALYSIS

Contents

Volume II: Exploratory and Confirmatory Factor Analysis

4. Exploratory Factor Analysis

EFA is a commonly used method for the elucidation of latent variables based on the re-dimensionalization of variables. Similar to but distinct from principal components analysis (PCA) (see Section 7), EFA effectively draws new axes over the existing data, creating some factors that capture as much as possible of the total variability in the variables while leaving other factors as residual error. The net result is that there are always as many factors produced as there are variables inputted into the procedure, but (ideally) a small number of factors can be used to represent the majority of the total variability expressed by the original variables. Ferguson and Cox (1993) provide a non-technical overview of EFA and provide a useful summary of advice based on the consensus of the methodological literature. Fabrigar et al. (1999) suggest five focal points to consider when using EFA in empirical research and illustrate each of these through a series of empirical examples. They also conduct a brief literature review that highlights poor research practice in the use of EFA. Hogarty et al. (2005) investigate minimum sample sizes for achieving high-quality solutions in EFA. They find that there is no single formula to determine a minimum, but that as expected data with higher communalities and fewer factors require less data for good estimation. In other words: the stronger the signal, the less data are needed to reliably identify it. This section concludes with three analytical reviews of the use of EFA in three different disciplines. In organizational behavior, Conway and Huffcutt (2003) cover much the same ground as Fabrigar et al. (1999) but come to a slightly more positive conclusion: using very similar criteria, they find that studies in which EFA is central to the research tend to use EFA more appropriately. In educational psychology, Henson and Roberts (2006) also find much poor technique, echoing both Conway and Huffcutt (2003) and Fabrigar et al. (1999) in their finding that many researchers simply use the default options on software packages instead of using appropriate methods. They suggest a seven-item checklist for improved research practice. Finally, Norris and Lecavalier (2010) review the use of factor analysis in developmental disability psychology. They find similar problems, including software package biases, and offer a table of recommendations for the use of EFA and the reporting of EFA results.

16

Exploratory Factor Analysis: A Users' Guide

Eamonn Ferguson and Tom Cox

Of the many and varied services offered by work and organizational psychologists, those associated with psychological testing are among the most frequent. In the development, interpretation and validation of tests, such psychologists often have to explain or predict behaviour in terms of constructs which are not directly observable. Such constructs are variously referred to as hypothetical or latent constructs. The most common approach to identifying and measuring such constructs in relation to psychological testing has been the application of factor analysis to self-report or behavioural data, although other techniques have been used (see: Dane, 1990; Kline, 1986). There is, therefore, a need to ensure that factor analysis is understood by work and organizational psychologists and applied properly and scientifically (Cattell, 1978; Comrey, 1978). With correct application, and the demonstration of adequate validity for any derived measurement scales, the frequently cited criticism of factor analysis as 'garbage in garbage out' can be refuted.

This paper describes the factor analytic process and provides heuristics for the successful application of the technique. These heuristics should be taken as guide-lines and not absolute recommendations where empirical evidence on their effectiveness is *not* available.

Although this paper focuses on exploratory factor analytic (EFA) procedures, much of what is discussed is equally applicable to confirmatory factor analysis (CFA). EFA and CFA can be distinguished at both a statistical and a

Source: *International Journal of Selection and Assessment,* 1(2) (1993): 84–94.

methodological level. The statistical differences are complex and beyond the scope of this discussion, but the interested reader is referred to Maxwell (1977) and Joreskog (1979b). Methodologically, CFA allows the researcher to specify exactly the nature of the factor model they are interested in, whereas this is not the case with EFA. For example, if a group of researchers were interested in developing a measure of work commitment but had only a tentative theory as to the structure of their commitment measure (i.e. how many factors it contained and which variables loaded on which factors) they would perform an EFA as described here. On the other hand, if they already had a strong theoretical model then CFA could be used to test their theory against other competing models.

Exploratory Factor Analysis

The basic factor model assumes that the observed variables reflect linear combinations of underlying factors. It is these *factors* which are causal in creating the derived factor structure. They can take two forms. First, there are *common factors*: those which are common to two or more observed variables but which can affect all observed variables. These can either be correlated or uncorrelated. Second, there are *unique factors* which are specific to each variable and orthogonal (i.e. statistically uncorrelated) to each other and to all common factors. The major goal of EFA is the identification of the minimum number of common factors required to reproduce the initial *correlation* or *covariance* matrix. As such, factor analysis is different from principal components analysis. Principal components analysis identifies components on the basis of variance; that is, the first principal component accounts for the most variance, then the next and so on. Factor analysis, on the other hand, tries to explain the set of correlations or covariances represented in the data. Factor analysis is thus concerned with *covariance* and is distinct from principal components analysis which is concerned with *variance* (see Maxwell, 1977; Joreskog 1979a; Tabachnick and Fidell, 1989).

To ensure that EFA is used in a *scientific* manner (see Cattell, 1978), a theoretically driven factor structure should be proposed prior to the analysis (Comrey, 1978). This allows for a degree of hypothesis testing to be undertaken: how similar is the emergent structure to the one proposed? A simple percentage agreement, or a hit score of how many variables load on the factors they were supposed to load on, may be used as a rough indicator of support for the original hypothesis. At a conceptual level this may appear to blur the distinction between EFA and CFA. While EFA can be used to test out hypothetical models at a fairly crude level, CFA is used as an exact test of new data against established models. However, CFA procedures allow for a degree of modification to any model, using their so-called modification indices, and when such procedures are applied CFA loses something of its confirmatory nature (Byrne, 1989).

EFA is also applied to data sets with no *a priori* expectations or simply as an exercise in data reduction. Although the former may contribute to hypothesis generation, some writers have criticised such practices (e.g. Comrey, 1978).

The EFA Process

The EFA process is completed in three stages: pre-analysis checks, extraction and rotation. Each of these stages is described below and the necessary decisions required at each stage are discussed.

Stage 1: Pre-Analysis Checks

The purpose of the pre-analysis checks is to ensure that: (1) a stable population factor structure can emerge from the sample; (2) items are properly scaled and free from biases, and (3) the data set is appropriate for the application of EFA. This pre-analysis stage is one of the most important, and yet it is the one which is most often overlooked.

Stable Factor Structure

Statisticians have suggested four heuristics for ensuring stable factor structures: (1) a minimum sample size (N), (2) a minimum ratio of subjects to variables, the N/p ratio (where p is the number of variables), (3) a minimum ratio of subjects to expected factors, the N/m ratio (where m is the number of expected factors) and (4) a minimum ratio of variables to expected factors, the p/m ratio (see Table 1). The relative merits of these heuristics have been discussed by Guadagnoli and Velicer (1988) who concluded that sample size is the most important heuristic. However, they add that mean factor loadings for a factor (factor saturation) is also a critical parameter, and if four or more items load on each emergent factor ≥ 0.6, then N is less relevant. However, the importance of N increases when both the factor saturation and the ratio of variables to expected factors (p/m) are low. In this case it has been argued that an N of at least 300 is required (Guadagnoli and Velicer, 1988). However, other writers have suggested that smaller Ns might be acceptable (for example, Kline, 1986 suggests $N = 100$). Guadagnoli and

Table 1: The type of heuristic, its range and advocates for producing a stable factor structure

Rule	*Range*	*Advocate*
Subjects-to-variables ratio (N/p ratio)	between 2:1 and 10:1	Kline (1986); Gorsuch (1983); Nunnally (1978)
Absolute minimum number of subjects (N)	100 to 200	Kline (1986); Comrey (1978)
Relative proportions of: variables to expected factors (p/m ratio), and subjects to expected factors (N/m ratio)	between 2:1 and 6:1	Cattell (1978)

Velicer (1988) demonstrated that, at high saturation, increasing N beyond the necessary minimum is a waste of time. Contrary to accepted wisdom, they also argued that, if the number of variables was then increased, no further increase in the number of subjects would be required. Based on this work, it is suggested that researchers pay particular attention to sample size (N), rather than the N/p and p/m ratios, and ensure a minimum N of 100.

Guadagnoli and Velicer (1988) have used the parameters N and factor *saturation* to develop a coefficient which gives an *a-posteriori* measure of factor stability (equation 1). This coefficient, *the stability coefficient* (Y), indicates how stable a factor structure is relative to the population from which it was drawn. However, as yet no calibration exists for Y, the only recommendation being that the smaller it is, the more stable the factor structure.

Equation 1

$$Y = 1.10(Xa) - 0.12(Xb) + 0.066$$

where:

Y = the average distance between a population loading and a sample loading (the stability coefficient)
Xa = the reciprocal of the square root of N
Xb = the average loading on a factor (factor saturation)

The availability of this equation means that issues of sample size (N) become less important since stability can be measured *a-posteriori*. Thus researchers should attain samples of no less than 100 subjects but also check the stability of their final solution using the stability coefficient.

Sampling: EFA should be carried out on a random sample from the population to which it is going to be generalized, unless it can be shown that the factor structure does not change as a function of population (see discussion of factor congruence below).

Item Scaling and Bias

Although either true interval or ratio scales are ideal for EFA, these are rarely achieved in practice and Likert-type scales (e.g. five-point scales) are often deemed adequate in psychological investigations (Comrey, 1978). Where Likert-type scaling is inappropriate (for example, in the measurement or such dichotomies as biological sex, or with forced choice questions), it is acceptable to use EFA as long as the initial correlation procedure is based on the phi coefficient instead of the Pearson's product-moment correlation coefficient (Pearson's coefficient) (Comrey and Levonian, 1958; Parkes, 1985). Phi coefficients are approximations of Pearson's coefficient but tend to be slightly smaller in size. This does not mean that they are unreliable, but they can only be used when a dichotomous scoring scheme has been utilized, as with Rotter's (1966) Locus of Control scale.

Skew, Kurtosis and Multivariate Normality

EFA techniques require that the variables used demonstrate univariate normality: that is, it is assumed that each variable conforms to the normal distribution curve (when the mean is in the centre of the distribution). Confirmatory techniques require multivariate normality: that is, the sum of all the variables conforms to a normal curve (Breckler, 1990; Joreskog and Sorbom, 1984), The coefficients of skewness and kurtosis tell the researcher whether or not each variable shows univariate normality. Skewness describes the extent of symmetry of a distribution, and a skewed variable is one whose mean is not in the centre of the distribution. Variables can be positively skewed, indicated by a + associated with the coefficient, and this occurs when most of the contributing scores are low. Conversely, variables can be negatively skewed. The coefficient of skewness is zero when there is no skew at all. Kurtosis summarizes the extent of 'peakedness' of a distribution. If a distribution is sharply peaked a positive kurtosis is evident, and if it is negative it means that the distribution is too flat. Most computer packages, SPSS[x] included, return a value of zero if the distribution is not kurtotic.[1] When a variable is perfectly normally distributed, its skewness and kurtosis coefficients are both zero, however, in reality this is rarely achieved. At present no real guide-lines exist on how to deal with varying degrees of skew and kurtosis with regard to EFA. A simple heuristic is therefore presented below.

A Heuristic or Dealing with Skew and Kurtosis

Muthen (1989) and Muthen and Kaplan (1985) have recently discussed the effects of skew and kurtosis on the performance[2] of a number of factor estimators. From these studies, three parameters emerged as important determinants of the effects of skew and kurtosis: (1) the absolute magnitude of skew and/or kurtosis for each variable, (2) the number of variables affected by skew and kurtosis, and (3) the proportion of the initial correlations within specified ranges (<0.2 and >0.5).

Muthen and Kaplan (1985) have argued that some degree of univariate skew and kurtosis is acceptable, for the *majority* of variables, if neither coefficient exceed $+/-2.0$[3]. If, however, there are *many* low correlations (<0.2) in the initial correlation matrix then greater skew is acceptable. These authors did not define what exactly they meant by *many* and for the purposes of the heuristic developed here many is defined as 60% or more.

With respect to deviant skew and kurtosis, there are four possible scenarios. In the first scenario, both skew and kurtosis exceed the acceptable limits. In the second scenario, skew exceeds the limits for some variables but kurtosis is acceptable. In the third scenario, skew is acceptable but kurtosis exceeds the limits for some variables. In the final scenario both skew and kurtosis are acceptable for all variables. It is the first three scenarios which are of concern, and a number of corrective *options* are available.

1. leave the deviant items in, or
2. identify the most appropriate transform[4] option to reduce the skew and/or kurtosis for all the affected variables (see Rasmussen, 1989) or
3. apply Muthen's (1989) Tobit factor analysis[5] program designed for highly censored data, or
4. remove all the deviant variables (Gorsuch, 1983).

The decision heuristic proposed here is set within the classical test theory and is designed to retain the maximum breadth of sampled variables while minimizing the possibility of spurious results.

The heuristic is used in the following manner. Initially, the percentage of variables adversely affected by either skew and/or kurtosis is calculated, ≤25% being taken as the cut off point for acceptability. This percentage figure was chosen because it is believed that if only 1/4 of the variables were affected they would not adversely affect the final solution. Having calculated the percentage of items affected by skew and kurtosis then the percentage of correlations within each correlation range is calculated. If sixty per cent or more of the correlations are:

1. below 0.20, then regardless of the number of variables affected by skew and kurtosis, all the variables can remain in the analysis (Muthen and Kaplan, 1985).
2. in the range of 0.21 to 0.49, then the affected variables require either transformation or removal. The transform option will not reduce the number of variables, and is to be preferred if more than a quarter of variables are affected.
3. greater than 0.50 then it may not be possible to do anything, as the matrix may show singularity. A matrix is said to be singular when off-diagonal variables are perfectly correlated. In the case of singularity, individual item sampling adequacies should be examined to identify the deviant items. Individual sampling adequacies are based on the Kaiser-Meyer-Olkin (KMO) test as described below. Items that have an individual sampling adequacy of less than 0.5 should be removed. Once these items have been removed then the procedures as for (2) above should be followed.

Social Desirability Response Set

Ideally, scales should be free from contamination by social desirability responding (Campbell, 1960; Kline, 1986; Jackson, 1970). Social desirability response set is a bias produced when individuals try to present themselves in a good light rather than 'honestly'. Elimination of this bias adds to both the reliability and the validity of any instrument. Three procedures have been considered in the literature.

Paulus (1981) has described a procedure, principal factor deletion, for eliminating social desirability bias which may be applied to EFA techniques. In essence, the procedure interrupts the analysis and screens the principal components for social desirability response set, removing from the factor matrix the *factor* which shows such confounding. The commonalities for the remaining variables are adjusted to account for the loss of some variables, and the remaining variables rotated. Communality expresses the variance in an observed variable accounted for by the common factors (see Loehlin, 1987). However, the major weakness in principal factor deletion is that it proceeds to rotation without re-analysing the remaining data set, and this leaves in doubt the extent to which the factor structure would have changed with re-extraction.

An alternative method of reducing social desirability response set has been suggested in terms of item construction. Forced choice questions are often suggested as less prone to social desirability response set (Rotter, 1966). The available evidence tends to support this suggestion. Askanasy (1985) investigated Rotter's Locus of Control Scale in relation to social desirability responding, using the forced choice and Likert-type response formats, and found that the forced choice scale showed the least social desirability confounding. However, caution should be exercised here since the reduced correlations with social desirability for forced choice measures may be an artifact of restricted range. That is, compared to Likert-type measures, forced choice scales have a smaller possible range of scores, and such restriction in range is known to reduce correlations (see Cohen and Cohen, 1983).

A third procedure is preferred by the authors. Every variable prior to EFA should be correlated with a standard measure of social desirability (e.g. the Marlow–Crowne Social Desirability Scale: Crowne and Marlow, 1961). Any variable(s) which show a significant correlation with this measure of social desirability should then be removed, thus reducing the possibility of the resultant factor structure being a product of social desirability responding.

Appropriateness of the Correlation Matrix

The correlation matrix for EFA needs to meet certain psychometric requirements (Cyr and Atkinson, 1986; Dziuban and Shirkey, 1974 and Gorsuch, 1983). Minimally this involves showing that there is some systematic covariation among the variables under consideration. This is important, as it can be shown that EFA will produce a solution on a set of randomly produced variables. Dziuban and Shirkey (1974) have argued that if this requirement of demonstrable covariation is not met then the results are not interpretable. Following Dziuban and Shirkey (1974), at least two statistics should be examined. The first is the Kaiser–Meyer–Olkin (KMO) test of sampling adequacy which indicates whether the associations between the variables in the correlation matrix can be accounted for by a smaller set of factors (a minimum value of 0.5 is required; Dziuban and Shirkey, 1974). This statistic is affected favourably by: (1) increased sample size, (2) increased number of variables, (3) increased level of correlations in the initial correlation matrix and (4) a decrease in the effective number of factors. The second recommended test is the Bartlett test of sphericity (BS). This tests the null hypothesis that no relationships exist between any of the variables. A significant test statistic (based on Chi square) indicates that there are discoverable relationships in the data.

If satisfactory results are obtained with these two tests, it is possible to proceed to extraction with confidence that the matrix derived from the data is appropriate for factor analysis.

Summary of Stage 1

Successful completion of stage 1 should reassure the researcher that they have a correlation matrix that is appropriate for factor analysis (KMO $\geq$ 0.5 and a non-significant BS). Furthermore, they know that the matrix is derived from properly scaled variables which are free from social desirability response set, and from problems with skew and kurtosis.

Stage 2: Factor Extraction

This section presents a description of the extraction process. The purpose of extraction is to identify and retain those factors which are necessary to reproduce adequately the initial correlation matrix. Therefore, decisions pertaining to the exact number of factors to extract are considered at this stage. Six commonly used heuristics are available to aid the decision process, and some are known to be more accurate than others (Zwick and Velicer, 1986). These six heuristics and their adequacies are described below. Data exists on the accuracy of the first four heuristics described (K1, *Scree test*, *Minimum Average Partial* and *Parallel Analysis*) (Zwick and Velicer, 1986), although the remaining two (*Very Simple Structure* and *Maximum Likelihood* χ^2) have not been examined in this way. However, an understanding of all these techniques places the researcher in a good position to evaluate the quality of a piece of psychometric research.

The first heuristic, known as the *Kaiser 1* (K1) rule (the most widely used) says, extract as many factors as there are eigenvalues (or latent roots) greater than one. An eigenvalue gives an estimate of the amount of variance associated with any factor, so that the rule involves retaining those factors which account for above average variance. This heuristic is only applicable when the initial communalities are all at unity (a correlation matrix drives the analysis). Where the initial commonalities are reduced (based on the