Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions

Pere J. Ferrando[*] and Urbano Lorenzo-Seva

University Rovira i Virgili (Spain)

When multidimensional tests are analyzed, the item structures that are obtained by Exploratory Factor Analysis are usually rejected when tested by Confirmatory Factor Analysis. This paper analyzes some of the aspects of the problem by means of simulation studies, and proposes a procedure that may be useful for dealing with the problem. The procedure is illustrated by means of an empirical example.

Key words: Confirmatory Factor Analysis, Exploratory Factor Analysis, Dimensionality assessment, goodness of fit in structural equation models.

A pervasive problem in the structural analysis of items designed to measure personality, attitude, psychopathology and other clinical constructs is that structures which were obtained using Exploratory Factor Analysis (EFA) tend to be rejected when tested statistically using a Confirmatory Factor Analysis (CFA) model. A typical scenario of well-planned research is as follows. First, an EFA solution which is clear and replicable is obtained in a series of preliminary studies, and second, a CFA model based on this EFA solution is tested in a new sample with the result that the model fits very badly. However, the fit might also be bad when the CFA is fitted to the same sample in which the EFA seems to produce a good solution.

The problem described above is well known among practitioners and has provoked different reactions. On the one hand, some authors have proposed alternatives to the direct use of the CFA at the item level: for example, item parcels (Bagozzi and Heatherton, 1994; Floyd and Widaman, 1995), or older factor-analytic methods, such as Procrustrean rotations (McCrae, Zonderman, Costa, Bond and Paunonen, 1996) or the oblique multiple group method (Bernstein and Teng, 1989). On the other hand,

---

[*]Correspondence may be sent to: Pere Joan Ferrando. Department de Psicologia. Universitat Rovira i Virgili. Ctra de Valls s/n, 43007 – Tarragona (SPAIN). email: pjfp@fcep.urv.es

some practitioners have proposed models based on very few items because these models appear to be more likely to show an acceptable fit. In addition, sometimes there is the practice to discard items 'ad hoc' until this acceptable fit is reached. These latter procedures lack substantive and theoretical foundations, they are likely to capitalize on chance, and, therefore, they cannot be recommended.

The purpose of the present paper is two-fold. Firstly, some of the aspects of the problem mentioned above are analyzed in detail by means of simulation studies; and secondly, a procedure that can help to solve the problem is proposed and illustrated by means of an empirical example.

Review of published item CFA

The first step we took to analyze the problem was to review 51 CFA published studies, which analyzed personality, psychopathology and attitude items, in 8 journals (*Anxiety, Stress and Coping, Educational and Psychological Measurement, European Journal of Personality, Journal of Applied Psychology, Journal of Personality Assessment, Journal of Research in Personality, Personality and Individual Differences* and *Psychological Assessment*) for the period 1994-1998.

The main characteristics of the studies reviewed can be summarized as follows: (a) the number of factors ranged from 1 to 8 with a median of 3; (b) the sample sizes ranged from 57 to 2,026 participants with an average of 449; (c) the number of items ranged from 7 to 85 with a median of 22, and the ratio items per factor ranged from 3 to 21 with a mean of 8. Most of the studies were based on previous EFA solutions, but in general they used much fewer items than the EFA solutions in which they were based used.

All of the studies reviewed used item response formats between 3 and 7 points and the most frequent was the 5 point format which was used in 26 (51%) of the studies. Of the 51 studies, 24 (47.1%) analyzed a correlation matrix while 27 (52.9%) analyzed a covariance matrix; 44 studies (86%) used maximum likelihood (ML) estimation; four (7.8%) used robust ML estimation; and three (5.9%) used Asymptotically Distribution Free (ADF) estimation. Even though ML is a normal theory-based estimation procedure, none of the 44 studies that used ML reported that they had tested the normality assumption.

Because the reviewed studies tested models of different sizes, it was decided to use the ratio between the chi-square test statistic value and the degrees of freedom of the model (χ^2/df) as a common measure of fit (χ^2 and df were reported in all of the studies). χ^2/df ranged from 1.31 to 19.5 with a mean of 3.86.

To sum up, our examination of 51 published applications demonstrated the magnitude of the problem, because virtually none of them reached a good fit from a statistical point of view and in most cases the fit was also unacceptable according to measures of fit alternative to the χ^2 test.

Delimitation of the problem

To clearly delimitate the problem, in most cases we will use the distinction Unrestricted-Restricted FA (UFA-RFA) solution instead of the more usual Exploratory-Confirmatory. We have done this mainly for two reasons. First, in most FA applications there is no clear EFA- CFA distinction: rather they fall on a continuum running from exploration to confirmation (Mulaik, 1972). Second, the distinction EFA-CFA is more conceptual than model dependent; for example, a study using traditional FA in which the number of factors and the approximate structure are hypothesized in advance is more confirmatory than exploratory, while a study in which a poor fitting CFA is modified 'ad hoc' is more exploratory than confirmatory (see Bollen, 1989). In contrast, the distinction Unrestricted-Restricted is clear. An unrestricted solution does not restrict the factor space, so unrestricted solutions can be obtained by a rotation of an arbitrary orthogonal solution, and, for the same data, all the unrestricted solutions will yield the same fit. A restricted solution, on the other hand, imposes restrictions on the whole factor space and cannot be obtained by a rotation of an unrestricted solution (Jöreskog, 1969).

When testing a covariance structure model by means of a χ^2 test of fit (or any goodness-of-fit index derived from it), there are two general classes of assumptions underlying the estimation procedure: distributional and structural (Satorra, 1990). In a typical RFA model the structural assumptions which are tested are concerned with: (a) the appropriate number of factors, (b) the specific pattern of loadings of each observed variable on these factors and (c) the relations among the factors (i.e. uncorrelated or correlated factors). So, an RFA may not fit the data because the distributional assumptions are violated, the number of hypothesized factors is inappropriate, the relations among variables and factors are not correctly specified or more than one of these factors occurs simultaneously.

To concentrate on the main problem considered here, we shall start by assuming that the estimation procedure is correctly specified for the distribution of the data and the number of factors is correct. If this is indeed so, the lack of fit of the model must be due to the additional restrictions imposed on the structure by the RFA. That is to say, it is assumed that researchers, possibly by means of various preliminary studies, have identified the main traits that explain the observed item responses and, also, that they know what the main items are that define the different traits.

The UFA model in *m* factors for a given item y_j is:

$$y_j = \lambda_{j1}\theta_1 + \lambda_{j2}\theta_2 + \cdots + \lambda_{jm}\theta_m + \delta_j \qquad (1)$$

If a clear structure is obtained for this model, then each item will have a large loading in only one factor, and small or minor loadings in the remaining factors. This is the concept of simple structure (Thurstone, 1947; chapter 14), which give rise to the idea of factorial simplicity as stated by Kaiser (1974) and Bentler (1977).

When an RFA structure is prescribed based on such a clear structure, the usual practice is to set to zero the 'minor' loadings found in the unrestricted solution (typically those below 0.20 ,0.30 or even 0.40). The corresponding restricted model is thus given by (for example):

$$y_j = \lambda_{j1}\theta_1 + 0\theta_2 + \cdots + 0\theta_m + \delta_j \qquad (2)$$

Expression (2) is Jöreskog's (1971) model for several sets of congeneric test (item) scores, and it also corresponds to the maximum simplicity of a factorial solution in the sense of Kaiser (1974) and Bentler (1977). In model (2) it is hypothesized that the 'minor' loadings found in the unrestricted solutions are consistent with exact zeros in the population (McDonald, 1985). In other words, in the usual applications each item is supposed to be a factorially pure measure of a sole trait, in the sense that only this trait contribute to the variance of the item (Thurstone, 1947).

In general, model (2) is recognized to be the ideal model because it assigns meaning to the estimated traits in the most unambiguous fashion (Gerbing and Anderson, 1984). However, it is also recognized that this model might be unrealistic when applied at the item level. In the cognitive and ability domains, it is generally agreed that most items require more than one ability to arrive at a correct response, and multidimensional models which have items with substantial parameter values in different dimensions have been developed for this type of item (Reckase, 1997). Likewise, in the personality domain it has been argued that individual items tend to be factorially complex (Cattell, 1986; Comrey and Lee, 1992; Floyd and Widaman, 1995). This is not to say that almost pure measures of certain traits cannot be constructed: rather, that the assumption that all the items in a multidimensional questionnaire are pure measures of a single trait is, perhaps an unrealistic one.

If model (1) is correct for the data, and model (2) is fitted, a bad fit would be expected due to the errors of specification, which in this case would be errors of omission (significant loadings incorrectly omitted or fixed to zero). If the misspecifications are not too large and the model is

small, perhaps the fit might be still satisfactory. The bigger the misspecifications and the more of them are, the worse the expected fit is.

A second potential problem is the bias on the estimates of free parameters (factor loadings, interfactor correlations) due to misspecifications. If model (1) is correct and model (2) is imposed on the data, then some distortion must be introduced in the estimates to keep the fixed parameters equal to zero. This problem is mentioned in Comrey and Lee (1992). However, neither this nor the previous problem seem to have been systematically studied.

Review of simulation studies

The problems which arise from using categorical variables are not a direct objective of the present research. However, because the typical item response formats reviewed above are polytomous and because categorization always poses preliminary problems in the analysis (eg. Bernstein and Teng, 1989), we will first consider this issue.

The conventional FA model is a model for continuous unlimited variables, and, strictly speaking, it should never be used with discrete variables. Simulation studies about this problem usually consider an FA model which holds for continuous variables, and study the robustness of the model when these variables are discretized (Bollen, 1989). If the continuous variables are normally distributed and the discrete variables are obtained by categorization at equal thresholds, then the distribution of the discrete variables will be approximately normal. In this case, if a normal theory based estimation procedure is used, according to the simulation studies, it is expected that: (a) the factor model will hold approximately for the discrete variables; (b) the loading estimates will be attenuated with respect to the true loadings of the continuous model, but the attenuation factor will be constant; and (c) the chi-square test statistic will be approximately correct. (Olsson, 1979; Boomsma, 1983; Muthén and Kaplan, 1985, 1992; Green, Akey, Fleming, Hershberger and Marquis, 1997).

We will turn now to the studies concerned with specification error. Problems which arise due to errors of omission have been studied considerably less than the problem discussed above. Kaplan (1988) used a full structural model and ML estimation. The measurement part of Kaplan's model was an oblique three-factor model with a total of 7 variables. The effects of the errors of omission were studied by fixing to zero a loading which was nonzero in the population. Kaplan found that the omission produced severe bias not only in certain measurement parameters, but also in certain structural parameters as well. The χ^2 test of fit, however, gave acceptable results in some cases.

Curran, West and Finch (1996) considered simultaneous violations of both the distributional and the structural assumptions in an oblique three-factor model with three indicators per factor. Their sample sizes ranged from 100 to 1000. Their model 3 was the most related to the present research and omitted two loadings which were nonzero (0.35) in the population. Under multivariate normality, they found that the ML χ^2 was correct for all sample sizes.

In conclusion, the studies reviewed give little information about the consequences of omitting multiple 'minor' but nonzero loadings for realistically sized models such as the ones mentioned above. On the one hand, Kaplan's results suggest that both the factor loadings and the interfactor correlations might be severely biased if a full information estimation procedure (such as the usual ML) is used. On the other hand, the χ^2 appear to be affected very little if one or two nonzero loadings are omitted in a small model, but whether it is affected or not under the conditions considered here is not known.

The present study

A series of Monte Carlo simulation studies were used to study the effects of model size and model specification on the assessment of fit and accuracy of the estimates.

Model specification

The general model considered was the multiple FA model corresponding to expressions (1) and (2). The well known covariance structure for this model is

$$\Sigma = \Lambda\Phi\Lambda'+\Psi \qquad (3)$$

In this study, Σ was a correlation matrix so that the unrestricted and restricted solutions could be better compared. Indeed, this is the common practice when EFA models are used, but it should not be considered so routinely in CFA applications. In the present case, all the CFA models considered met the following conditions: (a) Φ was a correlation matrix (i.e. diag(Φ)=I); (b) Ψ was unconstrained and (c) the only restrictions considered in Λ were fixed zeros. Under these conditions the RFA models were scale invariant (see Cudeck, 1989) and so the structure of the models was not modified with respect to the structure that would have been obtained by analyzing a covariance matrix. However, if other structural restrictions had been imposed (such as constraining two loadings or two error variances to be equal), the analysis of the correlation matrix would have been incorrect.

The restricted and unrestricted versions of model (3) were specified as follows. In the restricted version, each variable had a principal loading in only one factor and zero loadings in the remaining factors. In the unrestricted version, each variable had a principal loading in one of the factors and 'minor' loadings in the remaining factors.

Conditions. For both unrestricted and restricted models, two parameters were systematically varied: (a) model size and (b) interfactor correlations. Three conditions were considered for model size: small model (2 factors, 5 variables per factor), medium model (4 factors, 8 variables per factor) and large model (4 factors, 20 variables per factor). There were three conditions for the interfactor correlations: orthogonal model (zero interfactor correlations), low correlations (all correlations specified to be 0.15) and medium correlations (all correlations specified to be 0.30). All these conditions coincide with the empirical results reviewed above. Overall, the design has 2 (unrestricted-restricted) x 3 (size) x 3 (interfactor correlations)= 18 cells. The sample sizes (400) and the number of replications per cell were kept constant.

METHOD

For each of the 18 cells, 10 factor solutions were generated in the following way: (a) the primary loadings were selected at random from a uniform distribution between 0.5 and 0.8; (b) in the unrestricted solutions, the minor loadings were selected at random from a uniform distribution between 0.1 and 0.4 (in the restricted solutions the minor loadings were zero); (c) the signs of the loadings (both primary and minor) were also determined at random, so, in each of the solutions, approximately half of the loadings were positive and half negative. Once the factor solution had been generated, the corresponding correlation matrix was obtained by means of equation (3).

For each of the 10 factor solutions generated in each cell, 10 random samples (size 400) were generated from a population of standardized variables distributed in multivariate normal form. The correlation matrix was obtained from the factor solution (ie. the factor model fitted exactly in the population). Then the variables were discretized at four equal thresholds and used to compute the sample correlation matrices. So we analyzed 5-point Likert variables because they are the most frequent in applied research. However, the simulated variables are 'ideal' in the sense that they were all discretized at the same equal thresholds.

All of the 1,800 correlation matrices were analyzed with a restricted and an unrestricted model, so there was a total of 3,600 factor analyses. In

all cases the restricted models specified the primary loadings to be free parameters and fixed the minor loadings to zero. So, the restricted models were (approximately) correct when the population model was restricted, and were misspecified when the population model was unrestricted. The misspecifications were the nonzero minor loadings fixed to zero. As for the unrestricted models, which were (approximately) correct in all cases, first an arbitrary orthogonal solution was obtained, and then the patterns were obliquely rotated using direct Oblimin with $\gamma = 0$ (Clarkson and Jennrich, 1988).

The method of estimation was ML, chosen because it is the most widely used in the applications reviewed. ML was assumed to be appropriate in the present case because its distributional assumptions were met for the original variables and approximately met when these variables were discretized. Given that the RFA models were scale invariant, the ML goodness-of-fit test statistic was correct in all models, and lead to the same values that would have been obtained if covariance matrices had been analyzed. Indeed the test statistics were also correct for the unrestricted models (see eg. Swaminathan and Algina, 1978).

All analyses were conducted with the PC versions of PRELIS 2 and LISREL 8 (Jöreskog and Sörbom, 1996), and version 4.0 of MATLAB (1984) for a Silicon Graphics Computer. Details regarding data generation and other procedures are given in the appendix I.

Measures. The main index of fit considered in this paper is the χ^2 test. However, because we are comparing models in which the associated *df* vary substantially, the χ^2 should also differ substantially from model to model and it is not a useful index for comparisons. Therefore, we chose the χ^2/df ratio. For the most correctly specified models, the χ^2/df ratio was expected not to be too different from 1.0.

Even though the purpose of the present research was not to evaluate different fit indices per se, the applications reviewed above tend to rely more and more on fit indices other than the χ^2 test. So, we included two more indices for the reasons given below.

The χ^2 tests the null hypothesis that the model fits exactly in the population. This hypothesis is false for all the models considered here. In the most correct of the models (eg. a restricted solution analyzed using a restricted model), only the continuous variable solution holds exactly in the population; the discrete variable solution that is analyzed is expected to hold only approximately. To estimate the approximate fit of the models in the population, we used the Root Mean Square Error of Approximation (RMSEA), a measure of the discrepancy due to approximation per degree of

freedom. RMSEA is bounded below by zero, and is zero only if the model fits exactly in the population. Browne and Cudeck (1993) suggested that an RMSEA of about .05 or less means that the model has a close fit. More recently Hu and Bentler (1999) suggested a cutoff value of about .06. Details about RMSEA can be found in Browne and Cudeck (1993).

The third index considered is the Non-Normed Fit Index (NNFI), an incremental fit index which measures how well the model fits relative to a null model of uncorrelated observed variables. This particular index was chosen because it was first introduced by Tucker and Lewis (1973) in the context of factor analysis models, and it appears to be the only alternative fit index that is used in EFA applications. For example, the SAS FACTOR PROCEDURE (SAS Institute, 1992) output gives the NNFI under their original label of the Tucker-Lewis reliability coefficient. Hu and Bentler (1999) suggest an NNFI cutoff value of .95 as indicative of a good fit. Details about the NNFI can be found in Marsh, Balla and Hau (1996).

The impact of misspecification on the parameter estimates was assessed using two measures of accuracy: an individual measure and an omnibus measure. The individual measure was the bias and relative bias of parameter estimates, evaluated using

$$RB = \frac{\hat{\omega} - \omega}{\omega} \times 100 \qquad (4)$$

where $\hat{\omega}$ is the mean of the parameter estimates obtained across replications and ω is the population parameter value. For each cell, only the averages of the RB across the considered parameters will be reported. The omnibus measure was Tucker's (1951) congruence coefficient between each solution and the population pattern. The coefficient was computed using only the major loadings.

RESULTS

Figures 1, 2 and 3 show how the different indices of fit behaved. In each figure, the left hand side shows the average fit across the different interfactor correlation matrices. The right hand side shows the average fit across the different model sizes.

The general results are quite clear. When the model was (approximately) correct, the fit was excellent across all the conditions and for all of the indices. So, the UFA fitted well in all cases and the RFA fitted equally well when the solutions which were analyzed were restricted. As expected, these results suggest that when the model is correct, the categorization at equal thresholds does not appreciably affect the fit.

When the fitted model was not correct, i.e. when a restricted model was fitted to an unrestricted solution, the fit was unacceptable according to all the indices considered in the study. In the best case, the χ^2 was four times larger than the degrees of freedom, the RMSEA was above 0.10 and the NNFI was below .80. As for differential effects, on the one hand the fit appeared to be consistently worse when ϕ was orthogonal. On the other hand, the fit tended to get worse when going from the small to the medium models but not when going from the medium to the large models. In our opinion this last result could be due to a 'ceiling' effect.
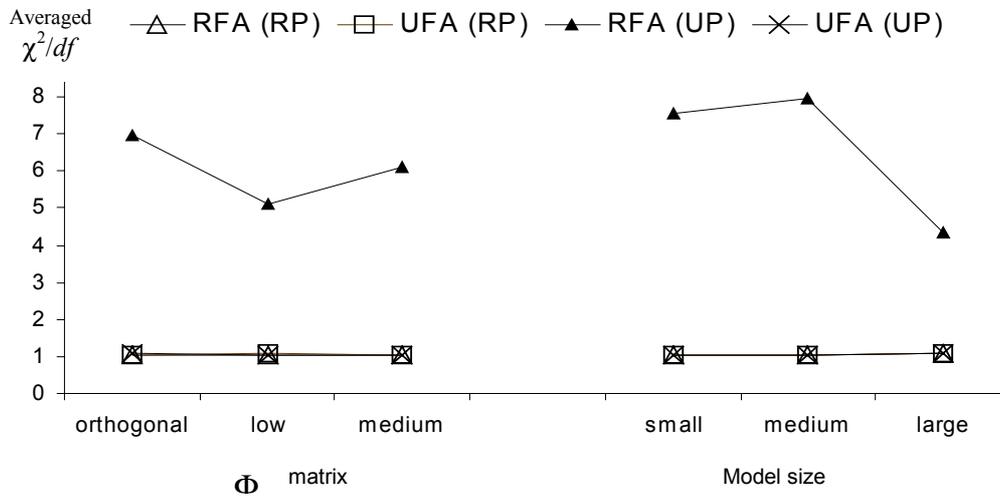

Figure 1. Averaged χ^2/df values for restricted (RFA) and unrestricted (UFA) models estimated when the populations models are restricted (RP) and unrestricted (UP) across the different conditions in the simulation study.
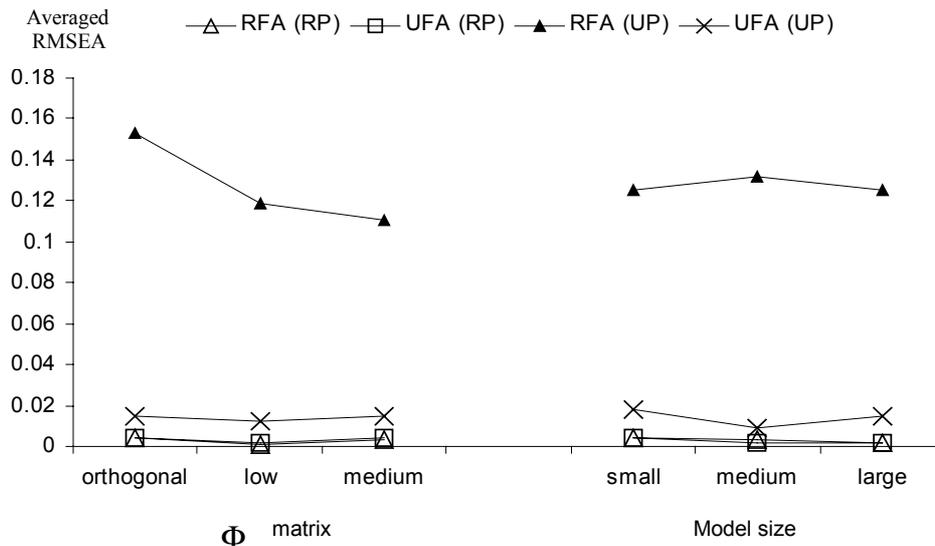
Figure 2. Averaged RMSEA values for restricted (RFA) and unrestricted (UFA) models estimated when the populations models are restricted (RP) and unrestricted (UP) across the different conditions in the simulation study.
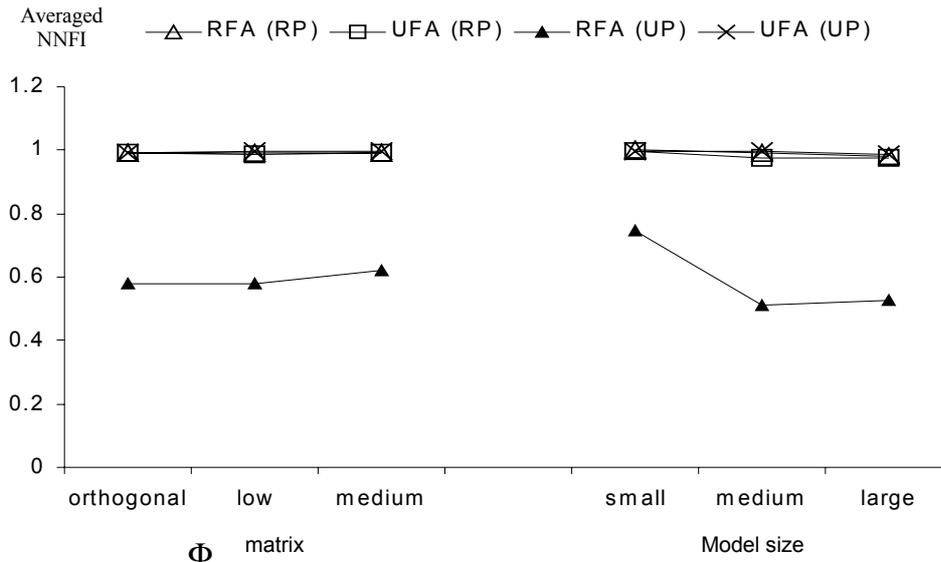
Figure 3. Averaged NNFI values for restricted (RFA) and unrestricted (UFA) models estimated when the populations models are restricted (RP) and unrestricted (UP) across the different conditions in the simulation study.

The results of the fit for the simulated RFA were worse, on average, than the results in the literature. In our opinion, this indicates that the RFA applications reviewed were in general more correct for the data than the models used here. In other words, most of the items in the applications may have been factorially purer than the ones considered here. Despite the results, many of the studies reviewed considered the model to be unacceptable and proposed models based on more factors. However, an unrestricted model with the same number of factors such as the one which is proposed in the next section may have given an acceptable fit in most cases.

Tables 1 and 2 show the accuracy in the parameter estimates when the restricted model was the correct population model (Table 1) and when the correct model was the unrestricted one (Table 2).

The results in Table 1 are quite clear. Both UFA and RFA closely approximated the population solutions. RB were small and congruence was almost perfect for all pattern sizes and inter-factor correlations. It should be pointed out that the pattern loading biases were all negative in sign. This

reflects the effect of attenuation due to the fact that the original variables had been categorized. However, the small RB and the values of the Tucker index near to one suggest that the attenuation factor was small and constant as expected.

The results in Table 2 are quite different. The pattern loading biases were larger than those of Table 1 for both UFA and RFA. In medium and large patterns, the RB tended to be larger for UFA than for RFA, which, at first sight, is an unexpected result given that UFA is more correct for the data. This may be because the RFA loadings were slightly overestimated due to the zero restrictions and this effect compensated the attenuation effect due to categorization. In any case, the RB were still reasonably low for both UFA and RFA.

In contrast to the reasonable accuracy of the loading estimates, considerable biases were found in the interfactor correlation matrices, and these biases were systematically larger in the RFA solutions. In all conditions the UFA tended to underestimate the inter-factor correlations, whereas the RFA grossly overestimated them. Finally, it appears that the overestimation was larger when lower the interfactor correlations were.

To summarize, the study shows that if an unrestricted model is correct for the data, the use of a restricted model would result in an unacceptable fit and biased parameter estimates. The fit will be bad according to the different indices considered, and even in small models. However, this last result does not directly contradict the idea that small models are more likely to reach an acceptable fit. Small models in which most of the items are factorially pure, or nearly so, might give a marginally acceptable fit.

As for the parameter estimates, the results show that even if the fitted restricted model is not correct for the data, the loading estimates might still be reasonably good. It appears that the main distortions produced by misspecification will be found in the interfactor correlation matrix in the form of gross overestimates of the correlations between factors.

A suggested procedure for testing an unrestricted model

Howe (1955) and Jöreskog (1979) gave conditions for uniqueness under factor rotation of a factor matrix Λ containing fixed zero elements. These conditions allow an unrestricted model to be specified as if it were a CFA model, so that any restricted model with fixed zeros is nested within the unrestricted model (i.e. the restricted model is formed by placing restrictions on free parameters in the unrestricted model). Tepper and Hoyle (1996) credit Stanley Mulaik in his capacity as Editor of Multivariate Behavioral Research for the idea of testing the unrestricted model in the

context of a nested sequence of models. The unrestricted model would be used as a baseline model for testing the hypothesis about the number of common factors.

Table 1. Accuracy of parameter estimates when the population models are restricted.

			Loadings			Inter-factor correlations	
		Bias	% Bias	Burt-Tucker		Bias	% Bias
Small Model	**Φ Orthogonal**						
	RFA	-0.0421	6.7603	0.9980		0.0631	
	UFA	-0.0419	6.7422	0.9979		0.0597	
	Φ low						
	RFA	-0.0333	6.9651	0.9981		-0.0028	13.2014
	UFA	-0.0333	6.9594	0.9981		-0.0094	15.1483
	Φ Medium						
	RFA	-0.0368	5.5227	0.9999		0.0024	9.7081
	UFA	-0.0381	5.7080	0.9998		-0.0117	9.8105

			Loadings			Inter-factor correlations	
		Bias	% Bias	Burt-Tucker		Bias	% Bias
Medium Model	**Φ Orthogonal**						
	RFA	-0.0370	5.6818	0.9999		0.0555	
	UFA	-0.0367	5.6381	0.9998		0.0497	
	Φ low						
	RFA	-0.0370	5.6811	0.9998		0.0002	18.6748
	UFA	-0.0386	5.9240	0.9998		-0.0182	26.5846
	Φ Medium						
	RFA	-0.0369	5.6296	0.9999		0.0003	9.5902
	UFA	-0.0430	6.5587	0.9998		-0.0324	21.6416

			Loadings			Inter-factor correlations	
		Bias	% Bias	Burt-Tucker		Bias	% Bias
Large Model	**Φ Orthogonal**						
	RFA	-0.0356	5.4652	0.9999		0.0423	
	UFA	-0.0357	5.4649	0.9999		0.0392	
	Φ low						
	RFA	-0.0364	5.5973	0.9999		0.0011	16.5969
	UFA	-0.0378	5.8034	0.9999		-0.0120	19.9753
	Φ Medium						
	RFA	-0.0363	5.6469	0.9999		-0.0016	9.7405
	UFA	-0.0422	6.5723	0.9999		-0.0323	21.6484

Table 2. Accuracy of parameter estimates when the population models are unrestricted

		Loadings			Inter-factor correlations	
		Bias	% Bias	Burt-Tucker	Bias	% Bias
Small Model	Φ Orthogonal					
	RFA	-0.0429	9.3815	0.9927	0.3692	
	UFA	-0.0429	7.6189	0.9966	0.1416	
	Φ low					
	RFA	-0.0465	11.6415	0.9898	0.1335	228.2194
	UFA	-0.0455	10.5899	0.9919	-0.0487	102.6353
	Φ Medium					
	RFA	-0.0628	22.6733	0.9692	0.2471	176.9209
	UFA	-0.0777	19.0389	0.9829	-0.1383	102.6171

		Loadings			Inter-factor correlations	
		Bias	% Bias	Burt-Tucker	Bias	% Bias
Medium Model	Φ Orthogonal					
	RFA	-0.0502	14.7792	0.9838	0.2623	
	UFA	-0.1011	20.5744	0.9760	0.0617	
	Φ low					
	RFA	-0.0608	19.8835	0.9721	0.1738	265.8566
	UFA	-0.1361	25.8235	0.9697	-0.0723	99.3512
	Φ Medium					
	RFA	-0.0811	28.1128	0.9419	0.1662	145.2563
	UFA	-0.1302	26.2235	0.9682	-0.1893	126.2342

		Loadings			Inter-factor correlations	
		Bias	% Bias	Burt-Tucker	Bias	% Bias
Large Model	Φ Orthogonal					
	RFA	-0.0404	9.4586	0.9944	0.1649	
	UFA	-0.1371	25.2854	0.9666	0.0404	
	Φ low					
	RFA	-0.0526	15.3560	0.9823	0.1343	215.7496
	UFA	-0.1307	24.6492	0.9623	-0.0896	119.4152
	Φ Medium					
	RFA	-0.0737	22.6780	0.9625	0.1501	114.4882
	UFA	-0.1368	24.5755	0.9728	-0.2017	134.4372

The testing of the unrestricted model can present two types of problem. First, the Howe-Jöreskog conditions are sufficient for the uniqueness under rotation, but not for identification as demonstrated by

Bollen and Jöreskog (1985) in a counterexample. Second, the procedure may led to convergence problems.

When the procedure is used only to test for the number of common factors the specified solution is arbitrary. However, here the procedure is also used to obtain a solution which is directly interpretable and which makes the subsequent rotations unnecessary. This is the use which was originally intended by Jöreskog (1969). Our experience suggests that, used in this way, the procedure hardly ever suffers from the aforementioned problems.

The specification of the unrestricted model is based on marker variables to give a recognizable factor. Marker variables are variables that previous research has shown to load highly on the factor being considered and very little on any other factor (Cattell, 1988; Gorsuch, 1974). Carroll (1978) points out that the 'ideal' design in FA should conform to a hypothesized structure consisting of at least three or four marker variables per factor, whereas Cattell (1988) considers two good marker variables per factor as a minimum. In the present case only one marker variable is used for each factor. These variables are the items which previous EFA-based research has shown to be the purest measures of their corresponding factors.

The unrestricted model is specified as follows:

1. For each of the common factors, one of the items (the marker) is specified to load only on that factor (i.e. the remaining loadings are fixed to zero). The remaining items are left free to load on every factor.

2. The variances of the common factors are fixed to unity.

If the model converges properly, it is equivalent to an EFA for the same number of factors. In other words, it would fit equivalently to any rotation of an EFA with the same number of common factors.

McDonald (1999, p. 181) describes an alternative procedure for specifying an UFA which is equivalent to an EFA. The procedure is standard in most structural equation programs and is based on an echelon pattern matrix in which zero factor loadings are imposed in the upper triangle of the factor loading matrix. The main difference between this method and the one proposed in our paper is that the constraints imposed in the echelon pattern are arbitrary and this fact, according to our experience using different datasets, causes the echelon method to be more prone to convergence problems.

An illustrative example

The smoking habits questionnaire (SHQ) was developed to assess the different situations that stimulated the desire to smoke in habitual smokers. The SHQ was designed to measure three general dimensions: (a) stress or

relief from stress; (b) activity or the commencement of activity; and (c) boredom. The items asked the subjects to imagine themselves in a given situation and to rate on a 5-point scale their desire to smoke. Nine items were intended to be measures of dimension (a), seven items measures of dimension (b) and six items measures of dimension (c). The item contents are shown in the appendix II.

A series of previous EFA in different samples resulted in a clear three dimensional structure in which the factors were defined by the hypothesized sets of nine, seven and six items. From previous EFA results, three items were selected as marker variables: item 1 (In the tense meetings) for dimension (a), item 16 (Before a very active task) for dimension (b) and item 20 (Watching a boring TV program) for dimension (c).

There were 255 participants for the present administration of the SHQ. First, a descriptive analysis was made: the departure of some item distributions from normality was found to be considerable and the absolute values of skewness and kurtosis were larger than one. The standardized multivariate coefficients of skewness and kurtosis were 47.5 and 20.3 respectively. Because the data was non-normal, normal-based procedures such as the usual ML may lead to distorted results. That is to say, the ML estimates will be consistent (if the model is correct) but the χ^2 test of fit and the standard errors of the estimates are likely be biased. Given that the sample was too small for A*DF* estimation, we opted for robust ML estimation which uses the (consistent) ML estimates obtained under the normality assumption. However the χ^2 and the standard errors are corrected for non-normality to provide a scaled χ^2 statistic (Satorra and Bentler, 1988) and robust standard errors (Chou, Bentler and Satorra, 1991). Robust ML estimation requires that the covariance matrix, not the correlation matrix, be analyzed.

Tables 3 and 4 show the parameter estimates, 95% confidence intervals (based on robust standard errors) and fit indices for the Unrestricted and the Restricted analysis of the data. The RMSEA and NNFI indices were computed based on the scaled χ^2.

At first sight, the differences between the UFA and RFA solutions appear not to be very large because the unrestricted solution is close to being a simple structure. The simplicity of the unrestricted solution can be assessed by inspecting the confidence intervals of the loadings: it appears that most of the secondary loadings are statistically non-significant. Even so, some of the trends which were found in the simulation study can also be found in this example. So, (a) when there are significant minor loadings in the UFA, which are fixed to zero in the RFA, the RFA estimates of the

principal loadings are larger than the UFA ones (see for example items 8, 11, 12, 15 18 and 22) and (b) the interfactor correlations are systematically higher in the RFA than in the UFA.

Table 3. Assessment of fit for the Unrestricted and Restricted Factor Analyses

χ^2	*df*	RMSEA	90% CI	NNFI
Unrestricted Factor Analysis				
333.29	168	0.062	(0.053;0.072)	0.912
Restricted Factor Analysis				
477.04	206	0.072	(0.064;0.080)	0.887

As for the assessment of fit in Table 3, the differences are important even if they are not too large. In general, the fit of the UFA model might be considered to be in the limit of acceptability: the NNFI is above .9; the RMSEA is about .06 and the χ^2/df ratio is 1.98, considerably below the average value found in the revised applications. On the other hand, the fit of the restricted model, although is not much worse, can no longer be considered acceptable. Because the models are nested, the χ^2 difference test can be used to assess whether the additional restrictions of the RFA (i.e. fixing the minor loadings to zero) significantly worsens the fit. The difference in chi squares is $\Delta\chi^2 = 143.75$ with 38 *df*. (p<.001). So, the fit is significantly worsened and at least some of the restrictions are not justified. It should be stressed that the differences between chi-squares are more important than the χ^2 values themselves. If the drop in χ^2 value is large compared to the difference in degrees of freedom the restricted model is inappropriate because some or all the restrictions are not justified. On the other hand, if the drop of χ^2 is close to the *df* difference, an unrestricted model may be able to increase the goodness of fit by capitalizing on chance.

The examination of the confidence intervals in the UFA solution might allow an intermediate restricted model to be specified. As one reviewer pointed out, this intermediate solution is related to the Independent-Clusters-Basis (ICB) model proposed by McDonald (1999). In an ICB model for correlated factors, at least two measures per factor are pure (three measures for uncorrelated factors).

As table 4 shows, the ICB model appears to be reasonable in the present case because we can find at least two measures per factor which have nonsignificant minor loadings.

As a reviewer suggested, the examination of the item contents is important; otherwise, the intermediate ICB model would be completely data-driven.

Table 4. Results of the unrestricted and restricted factor analysis on the SHQ data (95% confidence intervals underneath)

	Unrestricted Solution			Restricted Solution		
	F1	F2	F3	F1	F2	F3
I1	.587			.561		
	(.418;.756)					
I2	.682	-.131	.222	.744		
	(.351;1.012)	(-.381; .119)	(-.073; .517)			
I3	.722	-.167	.187	.718		
	(.396;1.048)	(-.413; .079)	(-.096; .470)			
I4	.966	-.105	.030	.880		
	(.586;1.346)	(-.458; .248)	(-.353; .414)			
I5	.661	.168	.126	.867		
	(.317;1.005)	(-.118; .454)	(-.210; .462)			
I6	.928	-.318	.281	.927		
	(.536;1.320)	(-.636; .001)	(-.096; .658)			
I7	.721	.007	.163	.852		
	(.400;1.043)	(-.272; .285)	(-.152; .477)			
I8	.737	.114	.314	1.086		
	(.407;1.068)	(-.190; .419)	(-.021; .649)			
I9	.623	.223	.148	.896		
	(.259; .987)	(-.073; .520)	(-.196; .491)			
I10	-.050	1.126	-.059		1.014	
	(-.412; .313)	(.875;1.377)	(-.371; .253)			
I11	-.093	.926	.421		1.144	
	(-.438; .252)	(.690;1.162)	(.129; .713)			
I12	-.106	.734	.400		.923	
	(-.449; .237)	(.479; .989)	(.104; .696)			
I13	-.212	1.008	.278		1.025	
	(-.530; .107)	(.775;1.241)	(-.053; .608)			
I14	-.012	1.151	-.076		1.048	
	(-.336; .312)	(.906;1.397)	(-.338; .186)			
I15	-.172	.641	.586		.894	
	(-.446; .101)	(.412; .869)	(.335; .838)			
I16		1.025			.998	
		(.881;1.168)				
I17	.219	-.199	.908			.971
	(-.072; .510)	(-.428; .031)	(.686;1.130)			
I18	.667	-.057	.371			.826
	(.343; .990)	(-.323; .208)	(.070; .673)			
I19	.083	-.351	1.282			1.142
	(-.261; .426)	(-.624;-.078)	(1.028;1.536)			
I20			1.102			1.085
			(.972;1.231)			
I21	.105	-.363	1.277			1.163
	(-.269; .478)	(-.628;-.098)	(.988;1.567)			
I22	-.105	.353	.677			.751
	(-.379; .169)	(.120; .587)	(.440; .914)			

	F1	F2	F3		F1	F2	F3
F1	1				1		
F2	.673	1			.696	1	
F3	.654	.525	1		.782	.590	1

Table 5. Results of the intermediate independent-clusters-basis model on the SHQ data.

	Restricted Solution		
	F1	**F2**	**F3**
I1	0.563		
I2	0.747		
I3	0.734		
I4	0.893		
I5	0.867		
I6	0.916		
I7	0.843		
I8	1.071		
I9	0.900		
I10		1.069	
I11		0.873	0.377
I12		0.677	0.350
I13		1.009	
I14		1.110	
I15		0.526	0.518
I16		1.042	
I17			0.949
I18	0.701		0.248
I19		-0.370	1.353
I20			1.092
I21		-0.379	1.365
I22			0.797
	F1	F2	F3
F1	1		
F2	.626	1	
F3	.783	.561	1

The inspection of table 4 points out that the following items are factorially complex: 11, 12, 15, 18, 19 and 21. Items 11 and 15 which were intended to measure 'Activity' load on the 'Activity' and 'Boredom' factors, and, effectively, their content may refer to both type of situations. Items 19 and 21 were designed as measures of 'Boredom' but they also load negatively in 'Activity'. This is also a plausible result, because both items refer to inactive situations. No clear substantive explanations were found for the complexity of items 12 and 18.

The ICB model was specified by constraining to be zero the minor loadings which were nonsignificant in the UFA solution. The specification and parameter estimates are shown in table 5.

The intermediate ICB model fitted the data reasonably well. The chi-square was 391.56 with 200 degrees of freedom. The point estimate of the

RMSEA was 0.061 and the NNFI estimate was 0.885. Indeed this model is also nested in the UFA, and the difference in chi-squares was $\Delta\chi^2=58.27$ with 32 df (p=0.0032), although significant, it appears that the fit does not get substantially worse when the additional restrictions are imposed, and the RMSEA indicates that the relative fit of the ICB model is even better than the fit of the UFA solution.

The ICB solution is reasonable and more parsimonious than the UFA solution. However the intermediate ICB model is an ad-hoc reespecification which is in part data-driven, and so the issue of capitalisation on chance would have to be addressed somehow somehow (MacCallum, Roznowski and Necowitz, 1992) specially in the present case in which the sample is quite small. If the intermediate ICB model were to be proposed evidence that it generalises beyond the used sample should be provided by cross-validation analysis. Also the causes of the factorial complexity of items 12 and 18 should be investigated.

DISCUSSION

The simulation study developed in this article is based on a set of items which are not all of them factorially pure and which are analyzed using a typical CFA model. This is the most usual situation in applied research. There are three important conclusions: (a) The CFA solution is expected to fit badly, even if the structure of the previous EFA is clear, for both big and small models; (b) the parameter estimates of the loadings that are not fixed to zero are expected to be reasonably accurate although slightly upwardly biased; and (c) the inter-factor correlations are expected to be grossly overestimated. To sum up, the results suggest that a CFA solution based on a correct EFA solution is likely to produce acceptable loading estimates, inflated interfactor correlations, and an unacceptable fit. As a result, the model will be rejected, and, as the literature shows, an (incorrect) CFA model based on a larger number of factors would probably be proposed.

In this article, we consider that the UFA model is better suited than the RFA model to the test items in most applications. Therefore, we propose a UFA model which is specified as a CFA model and equivalent to an EFA model for the same number of factors. This model can be used as a baseline model to test for the number of common factors in the way proposed by Mulaik (Tepper and Hoyle 1996), and also to give a meaningful solution which makes further rotations unnecessary. The χ^2 difference test between

the RFA and UFA, which are nested models, can be used to test the hypothesis that the additional RFA restrictions are justified.

Because modern CFA is firmly embedded in structural equation modeling, the specification of the UFA as a CFA model has several advantages, some of which have been exploited in the empirical example. A UFA specified as a CFA: (a) can be used in a nested sequence of models; (b) allows for correlated measurement errors among items; (c) can be estimated using methods for non-normal variables which are not available in the usual EFA programs and (d) its fit can be assessed using a variety of structural equation model indices which are also not available in standard EFA programs.

RESUMEN

El modelo factorial exploratorio y el modelo factorial confirmatorio en análisis de items. Si un test multidimensional se analiza mediante el modelo factorial exploratorio, la estructura obtenida suele ser rechazada cuando se somete a un análisis factorial confirmatorio. El presente trabajo analiza algunos aspectos de este problema mediante estudios de simulación y propone un procedimiento para evaluar la dimensionalidad que puede ser útil en investigación aplicada. Este procedimiento se ilustra mediante un ejemplo.

Palabras clave: Análisis Factorial Confirmatorio, Análisis Factorial Exploratorio, Evaluación de la Dimensionalidad, Bondad de ajuste en modelos de Ecuaciones Estructurales.

REFERENCES

Bagozzi, R.P. and Heatherton, T.F. (1994). A general approach to representing multifaceted personality constructs: application to state self-esteem. *Structural Equation Modeling,* 1, 35-67.

Bentler, P.M. (1977). Factor simplicity index and transformations. *Psychometrika,* 42, 2, 277-295.

Bernstein, I.H. and Teng, G. (1989). Factoring items and factoring scales are different: spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*. 105, 3, 467-477.

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

Bollen, K.A. and Jöreskog, K.G. (1985). Uniqueness Does not imply identification. A note on confirmatory factor analysis. *Sociological Methods & Research*, 14, 155-163.

Boomsma, A. (1983). *On the robustness of Lisrel (maximum likelihood estimation) against small sample size and non-normality*. Doctoral dissertation, University of Groningen.

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

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

Cattell, R.B. (1986). The psychometric properties of tests: consistency, validity and efficiency. In R.B. Cattell and R.C. Johnson (eds.) *Functional Psychological Testing* (pp 54-78). New York: Brunner/Mazel.

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.

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

Comrey, A.L. and Lee, H.B. (1992). A first course in factor analysis. Hillsdale: LEA.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, 105, 317-327.

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

Chou, C.P.; Bentler, P.M. and Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, 44, 347-357.

Floyd, F.J. and Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286-299.

Gerbing, D.W. and Anderson, J.C. (1984). On the meaning of within factor correlated measurement errors. *Journal of Consumer Research*, 11, 572-580.

Gorsuch, R.L. (1974). *Factor analysis*. Philadelphia: Saunders.

Green, S.B.; Akey, T.M.; Fleming, K.K.; Hershberger, S.L. and Marquis, J.G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modeling*, 4, 108-120.

Howe, W.G. (1955). *Some contributions to factor analysis*. Report No. ORNL-1919. Oak Ridge: Oak Ridge National Laboratory.

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

Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.

Jöreskog, K.G. and Sörbom, D. (1996). *PRELIS 2: User's Reference Guide*. Chicago: Scientific Software.

Jöreskog, K.G. and Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software.

Jöreskog, K.G. (1979) Author's addendum to: a general approach to confirmatory factor analysis. In J. Madgison (ed.) *Advances in factor analysis and structural equation models* (pp. 40-43). Cambridge: Abt Books.

Jöreskog, K.G. (1969). A general approach to confirmatory factor analysis. *Psychometrika*, 34, 183-202.

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

Kaplan, D. (1988). The impact of specification error on the estimation, testing and improvement of structural equation models. *Multivariate Behavioral Research*, 23, 69-86.

MacCallum, R.C., Roznowski,M, & Necowitz, L.B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin*, 111, 490-504.

Marsh, H.W. ; Balla, J.R. and Hau, K. (1996). An evaluation of incremental fit indices: a clarification of mathematical and empirical properties. In G.A. Marcoulides and R.E. Schumacker (Eds.) *Advanced structural equation modeling* (pp 315-353). Hillsdale: LEA.

Matlab (1984). *The Math Works*, Inc. Natick.

McCrae, R.R.; Zonderman, A.B.; Costa, P.T.; Bond, M.H. and 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. (1999). *Test theory: a unified treatment*. Hillsdale: LEA.

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

Muthén, B. and Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.

Muthén, B. and Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: a note on the size of the model, *British Journal of Mathematical and Statistical Psychology*, 45, 19-30.

Olsson, U. (1979) On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research,* 14, 485-500.

Reckase, M.D. (1997) The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21, 25-36.

SAS Institute (1992). *Sas/Stat User's guide Vol 1*. Cary: Author.

Swaminathan, H. and Algina, J. (1978). Scale freenes in factor analysis. *Psychometrika*, 43, 581-583.

Satorra, A. (1990). Robustness issues in structural equation modeling: a review of recent developments. *Quality & Quantity*, 24, 367-386.

Satorra, A. and Bentler, P.M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the American Statistical Association*, 308-313.

Tepper, K. and Hoyle, R.H. (1996). Latent variable models of need for uniqueness. *Multivariate Behavioral Research*, 31, 467-493.

Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago. Univ. Chicago press.

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

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