Common factor analysis versus principal component analysis: a comparison of loadings by means of simulations

de Winter, JCF; Dodou, D

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Common Factor Analysis versus Principal Component Analysis: A Comparison of Loadings by Means of Simulations

Joost C. F. de Winter & Dimitra Dodou

Common Factor Analysis Versus Principal Component Analysis: A Comparison of Loadings by Means of Simulations

JOOST C. F. DE WINTER AND DIMITRA DODOU

Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands

Common factor analysis (CFA) and principal component analysis (PCA) are widely used multivariate techniques. Using simulations, we compared CFA with PCA loadings for distortions of a perfect cluster configuration. Results showed that nonzero PCA loadings were higher and more stable than nonzero CFA loadings. Compared to CFA loadings, PCA loadings correlated weakly with the true factor loadings for under-extraction, overextraction, and heterogeneous loadings within factors. The pattern of differences between CFA and PCA was consistent across sample sizes, levels of loadings, principal axis factoring versus maximum likelihood factor analysis, and blind versus target rotation.

1. Introduction

Before conducting an exploratory data analysis, a researcher has to decide upon the subjects and variables to include, the method of analysis, the method for determining the number of dimensions to retain, the rotation method, and the method for calculating scores (e.g., Costello and Osborne, 2005; DiStefano et al., 2009; Fabrigar et al., 1999; Floyd and Widaman, 1995; Grice, 2001; Parmet et al., 2010; Preacher et al., 2013; Schmidt, 2011).

Concerning the method of analysis, both common factor analysis (CFA) and principal component analysis (PCA) are used frequently, with PCA appearing to be the most popular. Reviews investigating statistical practices in psychological, educational, marketing, and organizational research journals have shown that 12%–34% of the exploratory data analyses use CFA and 40%–67% use PCA (Conway and Huffcutt, 2003; Fabrigar et al., 1999; Ford et al., 1986; Glass and Taylor, 1966; Henson and Roberts, 2006; Park, Dailey, and Lemus, 2002; Peterson, 2000; Russell, 2002). PCA is also frequently used in areas that involve the analysis of a large number of variables, such as chemistry, human genetic variation, and artificial intelligence (e.g., Kaplunovsky, 2005; Paschou et al., 2007; Reich et al., 2008).

The popularity of PCA may be because of its simplicity and it being computationally less intensive than CFA (although the latter argument is hardly defensible nowadays) or because PCA is the default option in statistical packages such as SPSS. Steiger (2004) stated that what practitioners do "is, unfortunately, determined to a considerable extent by what SPSS, SAS, and other manufacturers of general-purpose statistical software are willing to implement in their programs. For better or worse, statistical practice is software-driven" (p. 71). Similar were the remarks made by Gorsuch (2003): "My personal opinion is that the rationales for CA [component analysis] developed as a post hoc explanation because so many used a computer package which had 'Little Jiffy' as the default" (p. 156). The occasional similarities in results produced by the two methods have led some users to consider PCA as a good approximation of CFA, but such practices have been a point of discussion among methodologists.

1.1. *Differences in Assumptions between CFA and PCA*

There are some fundamental similarities and differences between CFA and PCA, see Mulaik (2010), Ogasawara (2000), Schönemann and Steiger (1978), Steiger (1994), Velicer and Jackson (1990a), and Widaman (2007) for previous accounts. CFA and PCA are similar in the sense that both methods are concerned with describing a set of p manifest variables in terms of f latent variables, with $f \leq p$.

The CFA model presumes that each manifest variable is a function of f common factors (i.e., latent variables that influence more than one manifest variable) and one unique factor (i.e., a latent variable that influences only one manifest variable). This prediction can be described as follows:

$$X = YL'_F + Z + e \tag{1}$$

Herein, X is a $N \times p$ matrix of standardized scores on p manifest variables, Y is a $N \times f$ matrix of standardized common factor scores, L_F is a $p \times f$ matrix of least squares multiple regression weights (the common factor pattern) for predicting the scores in X from those in Y, and Z is a $N \times p$ matrix of unique factors. The common and unique factors are uncorrelated ($Y'Z = 0$) and the unique factors for different manifest variables are uncorrelated as well ($\frac{Z'Z}{N} = U^2$, a $p \times p$ diagonal matrix of unique factor variances). Because the unique factors are uncorrelated with each other and with the common factors, it is possible to interpret Y as containing common factors that account for correlations among manifest variables. Let us also assume an orthogonal common factor model, meaning that the f factors in Y are uncorrelated among each other. Because of sampling error and model error (e.g., nonlinear relationships between manifest variables and common factors, or minor factors; MacCallum et al., 2001), the common factor model will almost never hold exactly, and therefore e ($N \times p$) is included as a residual matrix.

Equation (1) may be rewritten as follows:

$$S = \frac{X'X}{N} = L_F L'_F + U^2 + E \tag{2}$$

with S a $p \times p$ sample correlation matrix and E a $p \times p$ matrix of residual correlations. A factor analysis procedure attempts to keep E as small as possible by optimally choosing L_F and U^2. Estimation in CFA typically employs an iterative procedure, where L_F is calculated at each U^2 found by the algorithm. The iterative principal axis factoring procedure is a popular method that provides a result that is asymptotically equivalent to the unweighted

least squares discrepancy function. The squared multiple correlations can be used as initial communalities. Factor loadings are then calculated by decomposing the reduced correlation matrix:

$$VKV' = S - U^2 \tag{3}$$

with V being a $p \times p$ matrix of eigenvectors normalized to unit norm and K an $p \times p$ diagonal matrix containing the corresponding eigenvalues of the reduced correlation matrix ($S - U^2$). The principal-axis factor loadings are:

$$L_F = V_r K_r^{1/2} \tag{4}$$

with V_r a $p \times f$ matrix of the first f eigenvector columns and K_r the diagonal $f \times f$ matrix of the corresponding eigenvalues. The new unique variances per variable are calculated as 1 minus the sum of the squared factor loadings (i.e., 1 minus the level of the communalities). The iterative procedure stops when a stable solution or a maximum number of iterations has been reached. Although L_F and U^2 can usually be uniquely determined from a given S (or X), Y and Z cannot, because the number of unknowns (f common factors $+ p$ unique factors) exceeds the number of equations (p equations, one for each of the p manifest variables). This is known as factor indeterminacy (e.g., Grice, 2001; Maraun, 1996).

In PCA, each manifest variable is a linear function of principal components, with no separate representation of unique variance:

$$X = Y_C L_C' \tag{5}$$

with Y_C being a $N \times p$ matrix of standardized component scores and L_C a $p \times p$ matrix of component loadings. PCA is a linear orthogonal transformation of the variables in X to principal components in such a way that the first component has the largest possible variance, the second component has the largest possible variance of the remaining data, etc., with the total p components explaining 100% of the variance. S can be decomposed as follows:

$$WMW' = S \tag{6}$$

with W a $p \times p$ matrix of eigenvectors and M a $p \times p$ diagonal matrix of eigenvalues of S. The component scores Y_C are:

$$Y_C = XWM^{-1/2} \tag{7}$$

The $p \times p$ matrix of component loadings (i.e., the correlation coefficients between Y_C and X) is calculated as follows:

$$L_C = WM^{1/2} \tag{8}$$

Usually, only the first f components are retained. The retained components account for less than 100% of the variance in the data.

Summarizing, CFA and PCA describe the correlation matrix in a different manner:

$$S = L_F L_F' + U^2 + E = L_C L_C' = L_{Cr} L_{Cr}' + R \tag{9}$$

with L_{Cr} a $p \times f$ matrix of retained component loadings and R a $p \times p$ matrix of residual correlations. In PCA, the component scores are uniquely determined by the data in X, and R cannot be diagonal.
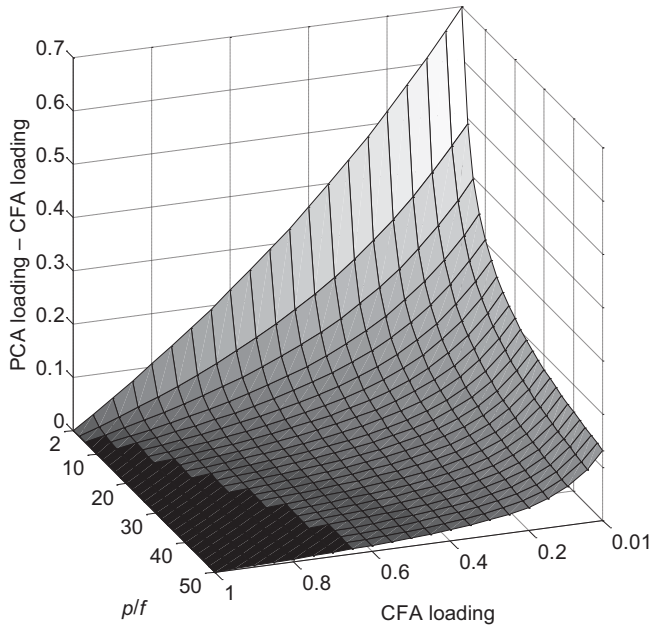
Figure 1. Difference between CFA and PCA loadings as a function of the number of variables per factor and the level of the CFA loadings.

A comparison of Eqs. (3) and (6) shows that the only difference between CFA and PCA is that CFA involves a reduction of the diagonal elements of the correlation matrix and PCA does not. CFA explains the off-diagonal elements of S according to a discrepancy function. In contrast, in PCA, the retained components optimally account for the 1s on the diagonal of S; that is, $L_{Cr}L'_{Cr}$ maximizes tr(S) among all possible orthogonal linear transformations of X (Jolliffe, 2002). PCA will often perform well in explaining the off-diagonal elements of S (Widaman, 2007), although this only occurs as a side effect. PCA's actual aim is to optimally account for the maximum amount of variance with the minimum number of components.

Note that the loadings L_F and L_{Cr} can be orthogonally rotated without compromising the validity of the above equations. The total variance accounted for by the f retained components remains the same when performing an orthogonal rotation, but the variance is distributed more evenly amongst the rotated components as compared to the unrotated components. Oblique rotation does compromise the variance maximization property of PCA. Nonetheless, researchers often rotate obliquely in order to obtain an interpretable pattern of loadings.

From Eq. (9), it can be inferred that the difference between CFA and PCA loadings will be small when the unique variances in U^2 are small. For loading matrices in which each manifest variable loads on only one factor, all factor loadings are equal, and no sampling error is present, the level of the component loadings l_C is always higher than the level of the factor loadings l_F, with the difference between l_C and l_F depending on the number of variables per factor p/f. This relationship is provided in Eq. (10) (adapted from Widaman,

2007; see also Fig. 1):

$$l_C = \sqrt{l_F^2 + \frac{1 - l_F^2}{(p/f)}} \tag{10}$$

Equation (10) shows that when the number of variables per factor is small, the PCA loadings will be high with respect to the CFA loadings. Equation (10) also shows that the difference between factor and component loadings becomes very small when the loadings reach unity or when the number of variables per factor grows to infinity. This latter point is in line with Velicer and Jackson (1990b): "Intuitively one can appreciate readily that the similarity will increase as the number of variables increases in view of the obvious fact that as the diagonal elements increase arithmetically the off-diagonal elements increase geometrically" (p. 104).

Several analytical and computational studies have shown that factor and component loadings diverge from each other when departing from sphericity (Gower, 1966; Rao, 1955; Sato 1990; Schneeweiss, 1997; Schneeweiss and Mathes, 1995; Widaman, 2007), where sphericity is defined as a condition in which unique variances are equal, and therefore—assuming there are no cross-loadings—loadings within factors are equal as well. The difference between CFA and PCA loadings is even larger when unequal loadings occur within correlated factors (Widaman, 2007).

A frequently heard rationale in favor of CFA is its generalizability (Steiger, 1994). If a large number of variables are explained by f common factors, then a subset of these variables can also be explained by the same common factors. This means that factor loadings *can* remain consistent for different subsets of variables. Widaman (2007) argued that "if one or more manifest variables are added to a battery of measures and any additional variables rely only on factors already well represented in the initial battery of p manifest variables (i.e., the additional variables rely on no new common factors), then all parameters associated with the original set of p manifest variables ... remain unchanged" (p. 190). The usefulness of components in generalizing to other sets of variables has been characterized as "awkward if not impossible" (Mulaik, 1990, p. 53), because components are linear combinations of manifest variables that "do not strictly represent inductive generalizations beyond data" (Mulaik, 1987, p. 299). PCA implicitly has no concept of a universe of variables to which it generalizes. Components describe the data at hand and any variable removed from or added to the dataset will lead to different component loadings.

1.2. *Previous Simulation and Empirical Studies Comparing CFA with PCA*

In possibly the first ever study comparing CFA with PCA, McCloy et al. (1938) used loading matrices with 10 variables and 3 rotated factors. The authors reported that PCA loadings were higher than CFA loadings, and that CFA loadings better corresponded to the population factor loadings than PCA loadings (see also Widaman, 2007).

Velicer (1974, 1976a, 1976b, 1977) and co-workers (Fava and Velicer, 1992a; Velicer and Fava, 1987, 1998; Velicer, Peacock, and Jackson, 1982) carried out a series of simulation studies to compare loading matrices produced by CFA (maximum likelihood factor analysis, specifically) and PCA, as a function of the level of loadings, number of variables, number of factors, and sample size. These authors mostly focused on clean pattern matrices (nearly equal loadings per factor, large number of variables per factor, and orthogonal factors), whereas Velicer et al. (1982) also tested structures with unequal loadings within factors and moderate (half the size of the main loadings) cross-loadings. They consistently found that the numerical differences between the loading estimates of the two methods were only

minor (e.g., average correlation of .98 between factor scores and corresponding component scores across all conditions tested in Fava and Velicer, 1992a). These results are supported by several empirical studies which analyzed data with both CFA and PCA and found no substantial differences in the loadings (e.g., Goldberg, 1990; Jensen and Weng, 1994). Jensen (1983), well known for his work on the general intelligence factor, noted: "To satisfy the purists who, from one theoretical standpoint insist on principal factor analysis, and to satisfy those who, for a different set of reasons (mostly mathematical) favor principal components analysis, I have often performed both types of analysis on the same data. Never is there more than an iota of difference in the results . . . or in any theoretical or practical conclusions one would draw from the analysis" (p. 314).

By reflecting on eight theoretical and practical issues (similarity of loadings, number of factors or components retained, effects of overextraction, improper solutions, computational efficiency, factor indeterminacy, latent vs. manifest variables, and exploratory vs. confirmatory analysis), Velicer and Jackson (1990a, 1990b) concluded that PCA is preferred over CFA based on the principle of parsimony: it is less computationally intensive, it does not generate improper solutions (i.e., Heywood cases), and the components are uniquely determined by the data. Velicer and Jackson's claims were heavily challenged and an issue of *Multivariate Behavioral Research* was devoted to the debate (Bentler and Kano, 1990; Bookstein, 1990; Gorsuch, 1990; Loehlin, 1990; McArdle, 1990; Mulaik, 1990; Rozeboom, 1990; Schönemann, 1990; Steiger, 1990; Velicer and Jackson, 1990a, 1990b; Widaman, 1990).

Snook and Gorsuch (1989) pointed out that Velicer and colleagues typically investigated clean and relatively large loading matrices: "The conclusion of Velicer et al. (1982) that 'the methods produce results which are equivalent' is not surprising given the fact that they used 36 variables" (p. 149). Furthermore, most of the studies by Velicer and colleagues used the difference between factor and component scores or the root mean squared error of the entire pattern matrix as similarity indexes, and did not investigate how individual loading coefficients differ between CFA and CFA. Snook and Gorsuch (1989) compared the accuracy of CFA (principal axis factoring, specifically) and PCA in reproducing population factor loadings. The authors found that CFA produced accurate estimates of the nonzero factor loadings, whereas component loadings were substantially higher than the factor loadings, especially when the number of variables was small. Furthermore, CFA appeared to provide more stable estimates of the zero loadings than PCA.

A limitation of the work of Snook and Gorsuch (1989) is that they only analyzed perfect cluster configurations (equal loadings per factor, no cross-loadings, orthogonal factors). Widaman (1990, 1993, 2007) extended the simulations of Snook and Gorsuch and confirmed that component loadings increase as the number of variables per factor decreases, whereas factor loadings are invariant to the number of variables per factor (see Eq. (10)). Widaman also showed that when factors are correlated and there is no sampling error, CFA produces accurate estimates of the interfactor correlations, whereas the higher loadings produced by PCA as compared to CFA lead to lower intercomponent than interfactor correlations. Widaman (1993, 2007) further investigated differences between CFA and PCA loadings in non-spherical cases, that is, unequal loadings within factors. The results showed that CFA provided accurate estimates of the factor loadings, whereas PCA loadings were more homogeneous than the CFA loadings. The differences between CFA and PCA grew when the factors were correlated (e.g., for a case with two factors, .8, .6, .4 loadings within factors, and interfactor correlation of .5, PCA's corresponding loadings were .77, .77, .71, and intercomponent correlation was only .32). In another simulation study, Beauducel

(2001) found that CFA was better able than PCA to recover weak factors; PCA tended to yield cross-loadings, dissolving the weak factor into the stronger factors.

The number of retained dimensions affects the two methods in a different manner. In PCA, if the number of retained components is increased or decreased, the previous unrotated component loadings will remain exactly the same. CFA, on the other hand, tries to fit a common factor model in an iterative manner. Hence, increasing or decreasing the number of factors will lead to different unrotated loadings. For CFA, at least two variables are needed to identify the loadings (for more precise information on identification criteria, see Anderson and Rubin, 1956), whereas PCA can produce single-variable components, or "singlets" as Beauducel (2001) called them. Practitioners may, therefore, be inclined to retain fewer factors than components (Jolliffe, 2002).

Fava and Velicer (1992b) found that overextraction (i.e., extracting more factors/components than the true/known number of factors) deteriorated the factor/component scores within each method, but even at the maximum level of overextraction the correlation *between* factor and component scores was high (.97 for 9 overextracted factors). Lawrence and Hancock (1999) found that, when overextracting, the PCA loadings were more distorted than CFA loadings with respect to their population counterparts. Based on simulation data, Fava and Velicer (1996) concluded that underextraction is a much more severe problem than overextraction, and that underextraction negatively affects factor analysis more than component analysis.

Although a number of simulation studies have previously compared factor with component loadings, the available evidence does not easily lend itself to comparative analysis. Velicer and colleagues investigated clean loading matrices with many variables and did not distinguish between the recovery of zero and nonzero loadings, but used crude similarity indexes instead. The studies by Widaman (1993, 2007), on the other hand, focused on a small number of variables and on severe distortions of a perfect cluster configuration. Furthermore, Widaman used population matrices, "an ideal (i.e., unreal) world in which the CFA model fits perfectly in the population" (Widaman, 2007, p. 189). In practice, the factor model does not perfectly fit in the population (due to model error) or in the sample (due to sampling error).

MacCallum and Tucker (1991), MacCallum et al. (1999), and MacCallum et al. (2001) illustrated that when unique variances are large (i.e., the loadings are low) the impact of sampling error is high. MacCallum and colleagues further illustrated how sampling error can degrade factor recovery as a function of factor loadings and number of variables. However, it is not yet clear how sampling error interacts with the difference between CFA and PCA loadings. Widaman (2007) argued that model misfit and sampling error "only obscure comparisons between CFA and PCA" (p. 189), suggesting that relative differences between CFA and PCA loadings will remain the same regardless of additional sources of error. Jolliffe (2002) hinted that there may be an interaction between CFA-PCA differences and sampling error: "In practice, the model itself is unknown and must be estimated from a data set. This allows more scope for divergence between the results from PCA and from factor analysis" (p. 161). Trendafilov et al. (2013) recommended that "PCA and EFA [exploratory factor analysis] solutions were to be directly compared according to their performance on any particular data set" (p. 210). Similarly, Rao (1996, see also Trendafilov et al., 2013) stated: "Some conditions under which the factor scores and principal components are close to each other have been given by Schneeweiss and Mathes (1995). It would be of interest to pursue such theoretical investigations and also examine in individual data sets the actual differences between principal components and factor scores" (p. 18).

Summarizing, although previous simulations have investigated parts of the spectrum of similarities and differences between CFA and PCA, a synthesis about differences in the loadings produced by both techniques is lacking. The remainder of this article is devoted to a computer simulation aiming to investigate differences between CFA loadings and PCA loadings in the case of sampling error. The aim of this work is *not* to designate which of the two methods *should* be used. The decision whether to use CFA or PCA has to be based on the purpose of the analysis, that is, explaining correlations in CFA versus data reduction in PCA. Rather, the aim of this work is to quantify the differences between the loadings produced by the two methods, and accordingly, the severity of error that occurs when using PCA while actually intending to explain a correlation matrix. Considering that many practitioners currently use both methods, it seems worthwhile to clarify the differences between CFA and PCA loadings.

We introduced various distortions of a perfect cluster configuration that have been investigated by others before (e.g., interfactor correlations, moderate cross-loadings, and unequal loadings within factors), while keeping the mean loadings constant in order to be able to make a valid comparison between the distortions. We also introduced some distortions which have not been investigated before (e.g., model error, nonloading variables, high cross-loadings, and under- and overextraction on unequal loadings within factors). Some of these distortions have been previously used by the authors for measuring the performance of CFA for small sample sizes (De Winter, Dodou, and Wieringa, 2009) and for comparing loadings obtained with principal axis factoring versus maximum likelihood factor analysis (De Winter and Dodou, 2012), but have not been used yet for comparing CFA and PCA.

2. Method

2.1. *Simulations*

Simulations were conducted for 20 true factor loading matrices: a baseline loading matrix of a perfect cluster configuration (level of loadings $\lambda = .6$, number of factors $f = 3$, number of variables $p = 18$) and 19 distortions of this baseline, for $N = 50,000$ and $N = 50$. The chosen baseline loading matrix is considered representative for the use of factor analysis in psychological research (Henson and Roberts, 2006). $N = 50,000$ was chosen to verify whether the loadings of each method converge for large N. $N = 50$ was considered a small sample size, introducing a considerable amount of sampling error and yielding marginally interpretable loadings (De Winter et al., 2009).

Based on the true factor loadings of each case, 10,000 $p \times N$ sample data matrices were generated for $N = 50,000$ and for $N = 50$, by using Hong's (1999) method. This method uses the Tucker et al. (1969) procedure for the generation of population matrices and on Kaiser and Dickman's (1962) algorithm for the generation of sample matrices. Hong's method allows for introducing model error and correlated factors in the correlation matrix. The correlation matrix of each sample was submitted to CFA and PCA.

For CFA, the method of principal axis factoring was used, as it generates fewer improper solutions than maximum likelihood factor analysis. The iterative procedure started with squared multiple correlations in the diagonal and was terminated when the maximum of the absolute differences of communalities between two subsequent iterations was smaller than 10^{-3}. The maximum number of iterations was set at 9999. A large number of iterations was chosen to ensure that all solutions converged to their asymptotic value. We also tried

a tolerance interval of 10^{-3} with 25 iterations as well as 10^{-6} with 9999 iterations. The results were hardly different. The former combination generated fewer improper solutions because (slowly) diverging solutions halted after 25 iterations. The latter combination slightly increased the number of improper solutions while requiring longer computational time. Note that from all the $N = 50$ analyses in this study, only 0.6% required more than 100 iterations.

Three factors were extracted in Cases 1–13, one factor in Cases 14 and 18, two factors in Case 15 (underextraction), four factors in Case 16, and five factors in Cases 17, 19, and 20 (overextraction). The loadings were Varimax rotated (normalized prior to rotation to have unit norm, and unnormalized after rotation), except for the oblique Cases 2, 9, and 20, in which direct quartimin (i.e., oblimin with gamma = 0; rotation algorithm created by Bernaards and Jennrich, 2005) was used instead. We used Varimax rotation because this is the most commonly used rotation in practice. Varimax is used in 51.7% of the exploratory data analyses, according to Henson and Roberts (2006).

To recover the order and sign of the loadings, Tucker's congruence coefficient (K; Tucker, 1951) was calculated for all $f \times f$ loading vector combinations between the sample loading matrix and the true loading matrix. Next, the reordering procedure of the sample loadings started with the highest absolute K and proceeded toward the lowest K until the sign and order of the factors were recovered.

If CFA yielded a communality higher than 1, then that solution was labeled as improper and disregarded together with the corresponding solution of PCA.

The following indices were calculated per sample loading matrix and subsequently averaged across all sample loading matrices:

- Mean of absolute nonzero loadings (*Mnz*). Nonzero loadings in the sample were defined as loadings corresponding to the nonzero true factor loadings.
- Mean of zero loadings (*Mz*). Zero loadings in the sample were defined as loadings corresponding to the zero true factor loadings.
- Correlation coefficient (CC) between the sample loadings and the true factor loadings. The *CC* was calculated between the sample loadings and the true factor loadings for each factor/component, and subsequently, the mean was taken across the factors/components. Note that in all cases examined, the true loadings on each factor consisted of zeros along with a number of relatively high (mean of .6) nonzero loadings. Hence, it can be expected that the *CC* values will be consistently quite high.
- Standard deviation of the nonzero loadings (*SDnz*). First the standard deviation of each individual loading coefficient of the pattern matrix was calculated across all repetitions. Next, the average was taken across the loading coefficients corresponding to the nonzero true loadings.
- Standard deviation of the zero loadings (*SDz*). *SDz* was calculated in the same manner as *SDnz*, with the average taken across the loading coefficients corresponding to the zero true loadings.

Several other indices were tried as well, including Tucker's congruence coefficient, root mean squared error, and score similarity (cf. Everett, 1983). These indices are composite measures of the level of the loadings and the variability of the loadings, and were therefore not used in the present study. A large root mean squared error, for example, could be either the result of consistent under/overestimation of the true loading coefficient (i.e., precise but inaccurate) or a consequence of a large variability of the obtained loadings around the true loadings (i.e., accurate but imprecise). The mean, standard deviation, and *CC* of loadings are a set of nonredundant indices. *Mnz* and *Mz* describe the level of the

loading independent of the variability of the loadings (note that we carefully avoid the term "bias"; for further discussion, see Velicer and Jackson, 1990b). The standard deviation is a measure of variability independent of the level of the loadings, and the *CC* describes the linear correspondence of the sample loadings and true factor loadings.

The following 20 cases were simulated:

Case 1: Baseline ($\lambda = .6, f = 3, p = 18$).

Case 2: Correlated factors. All three combinations of factors were substantially (.5) correlated.

Case 3: Random model error was introduced for every sample by means of 200 minor factors explaining 20% of the variance. The parameter determining the distribution of the minor factors was set at .25.

Case 4: Zero loadings replacing nonzero loadings. Six of the 18 nonzero loadings were set to zero.

Case 5: Reduced number of variables. The number of variables was reduced from 18 to 9.

Case 6: Increased number of variables to 36 from 18.

Case 7: Unequal loadings of the three factors. The first factor was loaded with .8, the second with .6, and the third with .4.

Case 8: Unequal loadings within factors. The nonzero loadings within each factor were alternated to .8/.4.

Case 9: Unequal loadings within correlated factors. This was the same as Case 8 but aggravated by correlating all three combinations of factors with .5.

Case 10: Cross-loadings (moderate). Four cross-loadings of moderate (.3) level were added. Two of the four cross-loadings were of opposite sign to prevent factor rotation. The main nonzero loadings were increased from .6 to .667, such that the mean of all true nonzero loadings remained equal to .6.

Case 11: Cross-loadings (high). Four cross-loadings of high (.5) level, two of which were of opposite sign, were added. The main nonzero loadings were increased from .6 to .622, such that the mean of all true nonzero loadings remained equal to .6.

Case 12: Unequal *p/f* (moderate). The third factor was weakened by decreasing the number of variables per factor from 6 to 3.

Case 13: Unequal *p/f* (severe). The third factor was weakened by including only 2 variables instead of 6.

Case 14: Underextraction (-2 factors). One factor/component was extracted instead of three. The nonzero loadings of the three factors were altered to .6/.5/.4. In this way, the nonzero loadings of the one retained factor were .6 and the other two factors were somewhat weaker.

Case 15: Underextraction (-1 factor). Two factors/components were extracted instead of three. The nonzero factor loadings of the three factors were altered to .7/.5/.3. In this way, the mean nonzero loadings of the first two retained factors were .6 and the third omitted factor was weak.

Case 16: Overextraction ($+1$ factor). Four factors/components were extracted instead of three.

Case 17: Overextraction ($+2$ factors). Five factors/components were extracted instead of three.

Case 18: Underextraction (-2 factors) on unequal loadings within factors. The nonzero loadings were altered to .8/.4 within the first factor, .7/.3 within the second factor, and .6/.2 within the third factor, and one factor/component was extracted instead of three.

Case 19: Overextraction (+2 factors) on unequal loadings within factors. The nonzero loadings within each factor were altered to .8/.4, and five factors/components were extracted instead of three.

Case 20: Overextraction (+2 factors) on unequal loadings within correlated factors. This was a combination of the distortions of Cases 9 and 19. We also tested the same case for underextraction (one factor instead of three) but did not include it in the analysis as the *CC* values of both CFA and PCA were too low (about .23) for $N = 50,000$ to be meaningful.

The mean of the nonzero true loadings was .6 in all 20 cases. Table 1 shows the factor loadings of the baseline and eight of the distortions. The analyses were conducted in MATLAB (Version R2012b, The MathWorks, Inc., Natick, MA).

2.2. *Supplementary Analyses*

$N = 50$ was used above as a relatively small sample size, in order to introduce a large amount of sampling error. Simulations were repeated with $N = 200$ to investigate loading estimates for conditions that may be more representative for empirical research. Simulations were also repeated with $\lambda = .45$ and with $\lambda = .75$ instead of .6, to investigate loading estimates for different levels of true loadings.

Among the available model fitting procedures for CFA, principal axis factoring and maximum likelihood factor analysis are the most widely used (Fabrigar et al., 1999). Hence, simulations were repeated with maximum likelihood factor analysis (MLFA) instead of principal axis factoring (PAF). For MLFA, a communality higher than .998 was labeled as improper, because the MLFA algorithm did not allow for communalities higher than 1.

Finally, we repeated the simulations by using Procrustes rotation to achieve the best least-squares transformation that fits the population pattern and to avoid rotational indeterminacy. Both oblique and orthogonal Procrustes rotations were performed, and the best fit with the true loading matrix in terms of Tucker's congruence coefficient was selected. Note that Procrustes rotation represents an idealized situation; the target pattern is usually not known when working with empirical data.

The MATLAB code for the simulations is provided as supplementary material.

3. Results

3.1. *Simulations*

3.1.1. *CFA versus PCA loadings with N = 50,000.* The mean nonzero factor loadings approximated the true factor loadings (.6) within .003 in all investigated cases. The mean nonzero component loadings were higher than the true factor loadings in all cases, with a maximum of .757 for a small number of variables (Case 5). For matrices in which each manifest variable loaded on one factor only and all true loadings on a factor were equal (Cases 1, 2, 4, 5, 6, 7, 12, and 13), the PCA loadings corresponded to Eq. (10) (Widaman, 2007). The mean interfactor correlations of CFA were .5 for Cases 2 and 9, and .497 for Case 20. The intercomponent correlations of PCA were .385 ($\approx .5 \cdot (.6/.683)^2$, cf. Widaman, 1993 for equations predicting intercomponent correlations), .437, and .329 for Cases 2, 9, and 20, respectively.

The *CC*s between the loadings estimated by both methods and the true factor loadings were above .99 in 19 of the 20 cases for CFA, and in 16 of the 20 cases for PCA. For CFA, a *CC* below .99 was found for model error (Case 3: .959), which is an expected result because model error represents lack of fit in the population. For PCA, a *CC* below .99 was

Table 1
True Factor Loading Matrices Used in the Simulation Study

Case 1 Baseline			Case 4 Zero loadings replacing nonzero loadings			Case 5 Reduced number of variables			Case 7 Unequal loadings between factors		
.6	0	0	.6	0	0	.6	0	0	.8	0	0
.6	0	0	.6	0	0	.6	0	0	.8	0	0
.6	0	0	.6	0	0	.6	0	0	.8	0	0
.6	0	0	.6	0	0	0	.6	0	.8	0	0
.6	0	0	0	.6	0	0	.6	0	.8	0	0
.6	0	0	0	.6	0	0	.6	0	.8	0	0
0	.6	0	0	.6	0	0	0	.6	0	.6	0
0	.6	0	0	.6	0	0	0	.6	0	.6	0
0	.6	0	0	0	.6	0	0	.6	0	.6	0
0	.6	0	0	0	.6				0	.6	0
0	.6	0	0	0	.6				0	.6	0
0	.6	0	0	0	.6				0	.6	0
0	0	.6	0	0	0				0	0	.4
0	0	.6	0	0	0				0	0	.4
0	0	.6	0	0	0				0	0	.4
0	0	.6	0	0	0				0	0	.4
0	0	.6	0	0	0				0	0	.4
0	0	.6	0	0	0				0	0	.4

Cases 8 and 9 Unequal loadings within factors			Case 10 Cross-loadings (moderate)			Case 12 Unequal p/f (moderate)			Case 18 Underextraction on unequal loadings within factors		
.8	0	0	.667	.3	0	.6	0	0	.8	0	0
.4	0	0	.667	−.3	0	.6	0	0	.4	0	0
.8	0	0	.667	.3	0	.6	0	0	.8	0	0
.4	0	0	.667	−.3	0	.6	0	0	.4	0	0
.8	0	0	.667	0	0	.6	0	0	.8	0	0
.4	0	0	.667	0	0	.6	0	0	.4	0	0
0	.8	0	0	.667	0	0	.6	0	0	.7	0
0	.4	0	0	.667	0	0	.6	0	0	.3	0
0	.8	0	0	.667	0	0	.6	0	0	.7	0
0	.4	0	0	.667	0	0	.6	0	0	.3	0
0	.8	0	0	.667	0	0	.6	0	0	.7	0
0	.4	0	0	.667	0	0	.6	0	0	.3	0
0	0	.8	0	0	.667	0	0	.6	0	0	.6
0	0	.4	0	0	.667	0	0	.6	0	0	.2
0	0	.8	0	0	.667	0	0	.6	0	0	.6
0	0	.4	0	0	.667				0	0	.2
0	0	.8	0	0	.667				0	0	.6
0	0	.4	0	0	.667				0	0	.2

Note. For Cases 2, 3, 16, and 17, the true loading matrix was identical to Case 1. Case 6 was as Case 1, but with 12 instead of 6 loading variables per factor. Case 13 was as Case 12, but with 2 instead of 3 loading variables on the third factor. Case 14 was as Case 1, but with .5 loadings on the second factor and .4 loadings on the third factor. Case 15 was as Case 1, but with .7 loadings on the first, .5 loadings on the second, and .3 loadings on the third factor. Cases 19 and 20 were as Cases 8/9.

found for model error (Case 3: .967), unequal loadings within correlated factors (Case 9: .980), and overextraction on unequal loadings within factors (Case 19: .989 and Case 20: .952).

Summarizing, the results at the population (i.e., $N = 50,000$) level showed that CFA loadings accurately corresponded to the true factor loadings, whereas PCA loadings were always higher than CFA loadings. The *CC*s of both methods were higher than .99 for the majority of the investigated cases. The full simulation results for $N = 50,000$ are provided in Table A1 of the supplementary material.

3.1.2. *CFA versus PCA loadings with $N = 50$.* CFA generated high rates of improper solutions for overextraction (Cases 16, 17, 19, and 20 with 9.0%, 25.7%, 26.9%, and 28.9%, respectively), small p (Case 5, 25.0%), and unequal p/f (Cases 12 and 13, with 4.0% and 18.2%, respectively). The remainder of the cases generated improper solutions for less than 2% of the samples (see Table A2 of the supplementary material).

Table 2 shows the simulation results for $N = 50$. The mean nonzero loadings (*Mnz*) of PCA were higher than the true factor loadings (.6) in most of the cases. The PCA loadings were highest when the number of nonzero loadings per factor was small (Case 5: .734, Case 12: .676, and Case 13: .670). The PCA loadings were lower than .6 when loadings were unequal within factors (Case 9: .597, Case 18: .599, and Case 20: .568). Overall, CFA loadings were slightly lower than the true factor loadings, with the lowest *Mnz* when factors were correlated (Case 2: .542, Case 9: .555, and Case 20: .521).

The factor correlations of both methods were substantially lower than the population factor correlations (.5). The mean observed correlations of CFA and PCA were .254 versus .202 for Case 2, .324 versus .235 for Case 9, and .302 versus .223 for Case 20.

The mean of the estimated zero loadings (*Mz*) was very close to .000 for all cases and for both CFA and PCA, except for correlated factors (Cases 2, 9, and 20). In these cases, CFA produced better estimates of the zero loadings than PCA.

The *CC*s of PCA were lower than those of CFA when the true factor loadings were unequal within factors and when over- and underextracting (Cases 4, 8, 9, 14–20). The difference between the *CC*s of CFA and PCA were largest for unequal loadings within correlated factors (Case 9: .856 vs. .792 for CFA vs. PCA). Large differences between the *CC*s of the two methods were also noted for underextraction on unequal loadings within factors (Case 18: .854 vs. .796 for CFA vs. PCA) and for overextraction on unequal loadings within correlated factors (Case 20: .857 vs. .815 for CFA vs. PCA). Interestingly, the *CC*s of both methods were higher for cross-loadings (Case 11: .928 vs. .933 for CFA vs. PCA) than for the baseline (Case 1: .914 vs. .924 for CFA vs. PCA), indicating that factors/components are better defined when cross-loadings are present.

The PCA nonzero loadings were more stable (i.e., had a lower *SDnz*) than the CFA nonzero loadings, except for unequal loadings within correlated factors (Case 9), overextraction (Cases 16, 17, 19, and 20), and underextraction on unequal loadings within factors (Case 18). The zero loading variables of PCA were less stable (i.e., had a higher *SDz*) than those of CFA in all cases (mean *SDz* across all cases: .151 vs. .176 for CFA and PCA, respectively).

Table 3 shows the mean and standard deviations of loadings across repetitions for Case 9 (unequal loadings within correlated factors) and Case 13 (severely unequal p/f). For Case 9, both CFA and PCA underestimated the high (.8) true loadings and overestimated the low (.4) true loadings, this homogenization being stronger in PCA. The underestimation of the high true loadings was stronger than the overestimation of the low true loadings, which also explains why the PCA loadings were lower than .6 for Case 9 (see Table 2).

Table 2
Simulation Study Results for $N = 50$

Case	Mnz		Mz		Mean CC			Mean SDnz		Mean SDz	
	CFA	PCA	CFA	PCA	CFA	PCA	PCA>CFA	CFA	PCA	CFA	PCA
1. Baseline	**.594**	.666	.000	.000	.914	**.924**	**.994**	.121	**.102**	**.136**	.149
2. Correlated factors	.542	**.614**	**.068**	.069	.813	**.829**	**.864**	.171	**.165**	**.162**	.183
3. Model error	**.576**	.639	.000	.000	.846	**.856**	**.884**	.180	**.173**	**.180**	.199
4. Zero loadings replacing nonzero loadings	**.573**	.645	.000	.000	**.819**	.800	.113	.163	**.153**	**.174**	.218
5. Reduced number of variables	**.606**	.734	.000	.000	.897	**.924**	.971	.160	**.108**	**.145**	.169
6. Increased number of variables	**.593**	.631	.000	.000	.923	**.927**	**.999**	.106	**.097**	**.130**	.136
7. Unequal loadings between factors	**.594**	.667	.000	.000	.873	**.894**	**.954**	.131	**.126**	**.137**	.148
8. Unequal loadings within factors	**.595**	.655	.000	.000	**.931**	.919	.007	.108	**.104**	**.130**	.148
9. Unequal loadings within correlated factors	.555	**.597**	**.049**	.062	**.856**	.792	.012	**.164**	.198	**.155**	.200
10. Cross-loadings (moderate)	**.594**	.643	−.001	−.001	.948	**.954**	**.997**	.105	**.091**	**.122**	.129
11. Cross-loadings (high)	**.586**	.636	.000	.000	.928	**.933**	**.896**	.130	**.119**	**.134**	.145
12. Unequal p/f (moderate)	**.594**	.676	−.001	**−.001**	.893	**.910**	**.965**	.132	**.109**	**.144**	.161
13. Unequal p/f (severe)	**.583**	.670	.001	**.001**	.842	**.875**	**.846**	.146	**.126**	**.163**	.182
14. Underextraction (−2 factors)	.565	**.612**	.000	.000	**.835**	.805	.111	.125	**.115**	**.218**	.275
15. Underextraction (−1 factor)	**.588**	.651	.000	.000	**.874**	.871	.373	.124	**.111**	**.161**	.192
16. Overextraction (+1 factor)	**.585**	.641	.001	.001	**.903**	.899	.454	**.142**	.155	**.134**	.149
17. Overextraction (+2 factors)	.573	**.614**	**.001**	.001	**.891**	.875	.186	**.162**	.191	**.131**	.147
18. Underextraction (−2 factors) On unequal loadings within factors	.567	**.599**	.000	.000	**.854**	.796	.015	**.118**	.119	**.214**	.280
19. Overextraction (+2 factors) On unequal loadings within factors	**.582**	.630	.000	.000	**.928**	.918	.102	**.124**	.128	**.123**	.137
20. Overextraction (+2 factors) On unequal loadings within correlated factors	.521	**.568**	**.047**	.056	**.857**	.815	.131	**.174**	.203	**.135**	.167

Note. Mnz = mean of nonzero loadings; Mz = mean of zero loadings, CC = correlation coefficient with true factor loadings; $SDnz$ = standard deviation of the estimated nonzero loadings minus the true nonzero loadings; SDz = standard deviation of the estimated zero loadings. The method with Mnz closer to .6, Mz closer to 0, higher CC, more than 50% of sample solutions yielding a higher CC, lower $SDnz$, and lower SDz compared to the other method is in boldface.

Table 3

Means and Standard Deviations of CFA and PCA Loadings Across Repetitions for Case 9 (Unequal Loadings Within Correlated Factors) and Case 13 (Severely Unequal *p/f*) for $N = 50$. Results are Based on 9899 and 8183 out of 10,000 Repetitions (101 and 1817 Improper Solutions Excluded)

Case 9

Loadings						Standard deviations					
CFA			PCA			CFA			PCA		
.725	.070	.070	.715	.103	.104	.142	.126	.125	.161	.156	.155
.379	.027	.027	.467	.017	.020	.186	.187	.185	.233	.244	.242
.724	.071	.069	.714	.104	.103	.145	.127	.126	.163	.158	.156
.377	.030	.027	.465	.023	.019	.185	.187	.185	.232	.243	.244
.722	.071	.070	.714	.104	.103	.145	.127	.124	.162	.157	.154
.379	.028	.027	.465	.020	.018	.185	.184	.185	.234	.242	.244
.071	.723	.070	.104	.714	.104	.127	.144	.126	.157	.162	.157
.026	.380	.031	.018	.465	.024	.186	.187	.186	.243	.234	.246
.070	.723	.072	.102	.713	.106	.126	.143	.127	.156	.162	.158
.032	.380	.025	.024	.468	.014	.184	.183	.184	.241	.230	.242
.069	.724	.071	.102	.715	.104	.126	.144	.128	.155	.161	.158
.026	.381	.029	.017	.468	.021	.184	.187	.186	.243	.234	.243
.070	.070	.723	.105	.106	.712	.125	.126	.140	.157	.159	.164
.026	.029	.379	.019	.022	.465	.184	.185	.184	.242	.243	.233
.070	.070	.724	.105	.105	.712	.124	.127	.141	.156	.159	.163
.029	.027	.381	.021	.020	.467	.181	.183	.184	.239	.243	.231
.071	.071	.723	.105	.106	.711	.125	.127	.141	.156	.160	.163
.031	.025	.381	.025	.018	.466	.186	.185	.186	.246	.241	.234

Case 13

Loadings						Standard deviation					
CFA			PCA			CFA			PCA		
.590	−.001	.001	.662	−.001	.001	.132	.136	.199	.119	.148	.226
.591	−.002	.005	.662	−.001	.003	.132	.137	.199	.118	.150	.226
.588	.000	.002	.660	.001	.001	.131	.135	.200	.118	.148	.227
.591	.000	.003	.663	.000	.002	.131	.136	.200	.117	.149	.225
.591	.002	.001	.663	.001	.002	.132	.134	.201	.119	.148	.226
.591	.002	−.001	.664	.002	−.001	.132	.136	.199	.116	.149	.222
.001	.591	.001	.001	.662	.000	.136	.132	.204	.148	.119	.231
−.001	.589	.003	−.001	.662	.001	.134	.133	.203	.147	.119	.228
.000	.590	.002	.000	.663	.001	.137	.130	.200	.148	.116	.225
.000	.590	.002	.000	.663	.002	.134	.132	.200	.147	.118	.225
.000	.590	.003	.000	.663	.003	.137	.131	.204	.151	.116	.228
.000	.590	.000	.000	.661	−.001	.136	.132	.201	.150	.120	.228
.002	.001	.534	.003	.001	.709	.134	.134	.233	.149	.148	.174
.000	.000	.534	.000	.000	.709	.133	.134	.233	.147	.149	.173

Table 4

Loading Estimates by PCA After Having Excluded the PCA Solutions Corresponding to Improper Solutions Generated by CFA Versus Loading Estimates by the PCA Solutions Corresponding to the Improper Solutions Generated by CFA. Number of Proper Solutions out of 10,000: Case 5: 7504; Case 13: 8183; Case 17: 7426

	PCA solutions corresponding to proper CFA solutions	PCA solutions corresponding to improper CFA solutions
Case 5		
Mnz	.734	.708
Mean *CC*	.924	.897
Mean *SDnz*	.108	.152
Mean *SDz*	.169	.184
Case 13		
Mnz	.670	.681
Mean *CC*	.875	.903
Mean *SDnz*	.126	.115
Mean *SDz*	.182	.162
Case 17		
Mnz	.614	.616
Mean *CC*	.875	.878
Mean *SDnz*	.191	.190
Mean *SDz*	.147	.146

Note. Mnz = mean of nonzero loadings; *CC* = correlation coefficient with true factor loadings; *SDnz* = standard deviation of the estimated nonzero loadings minus the true nonzero loadings; *SDz* = standard deviation of the estimated zero loadings. Mean of zero loadings (*Mz*) was .000 in all cases.

The PCA nonzero and zero loadings were less stable than those of CFA. For Case 13, the nonzero PCA loadings were higher than the true factor loadings, with this deviation from the true loadings being the greatest for the third factor/component, which had only two salient loadings. As Table 3 shows, the weak factor/component was also the least stable for both methods. The PCA nonzero loadings were more stable than the CFA nonzero loadings, whereas the zero loading variables of PCA were less stable than those of CFA (see also Table 2).

Table 4 shows the loading estimates by PCA, corresponding to proper CFA solutions and improper CFA solutions. Cases 5, 13, and 17 are shown, as these generated a high percentage of improper solutions. It can be seen that loading estimates by PCA were about the same, regardless of whether CFA provided a proper solution.

3.1.3. *Supplementary analyses.* Table A3 of the supplementary material contains the results of the simulations for $N = 200$ instead of $N = 50$. For both CFA and PCA, $N = 200$ yielded higher *CC*s than $N = 50$ (mean *CC* across all cases: .970 vs. .881 for CFA; .964 vs. .876 for PCA). *Mnz* of CFA were higher for $N = 200$ than for $N = 50$ (mean *Mnz* across all cases: .594 vs. .578). For PCA, the mean nonzero loadings (*Mnz*) were also higher for $N = 200$ than for $N = 50$ (mean *Mnz* across all cases: .669 vs. .639).The rates of improper CFA solutions were lower for $N = 200$ than for $N = 50$ (except for Case 13). Compared to

$N = 50$, *SDnz* and *SDz* were about halved, in line with the commonly known fact that standard errors decrease by the square root of sample size. The relative performance of CFA versus PCA (i.e., which method provides the higher *Mnz, CC, SDnz*, and *SDz* per case) was highly similar to $N = 50$.

Table A4 of supplementary material contains the simulation results when using $\lambda = .45$ instead of .6. As compared to $\lambda = .6$, this condition generated higher rates of improper solutions, reaching 57.0% for Case 5. Moreover, $\lambda = .45$ led to lower *CC*s than $\lambda = .6$, confirming earlier studies that a larger sample size is required when loadings are lower (e.g., MacCallum et al., 1999). Across all cases, CFA and PCA yielded a mean *CC* of only .710 and .708, respectively (cf. .881 and .876 for $\lambda = .6$). Both methods produced less stable solutions for $\lambda = .45$ than for $\lambda = .6$, as indicated by the higher *SDnz* (.203 vs. .139 for CFA, .221 vs. .135 for PCA, all four *SDnz* calculated across the 20 cases) and *SDnz* (.190 vs. .151 for CFA, .230 vs. .176 for PCA). Both methods yielded low inter-factor/intercomponent correlations in the oblique cases (about .17 and .12 for CFA and PCA, respectively, while the interfactor correlation in the population was .5). The relative performance of CFA versus PCA, in terms of which method provides higher *Mnz, CC, SDnz*, and *SDz*, was similar to $\lambda = .6$.

Table A5 of the supplementary material contains the simulation results for $\lambda = .75$ instead of .6. The relative performance of CFA and PCA was again very similar to $\lambda = .6$, but $\lambda = .75$ yielded higher *CC*s (mean *CC* across all cases: .952 vs. .881 for CFA, .944 vs. .876 for PCA) and more stable loadings than $\lambda = .6$.

Table A6 of the supplementary material shows the results of the simulations when using Procrustes instead of blind (i.e., Varimax or Oblimin) rotation. In all but one (Case 14) of the cases, *Mnz* were higher for Procrustes than for Varimax (differences ranging between .001 and .020 for both CFA and PCA). The relative performance of CFA versus PCA under Procrustes rotation was similar to the Varimax solution in terms of *CC*, with all *CC*s being slightly higher for Procrustes (differences between the two rotation methods ranging between .001 and .020 for both CFA and PCA). For the three oblique factor patterns (Cases 2, 9, and 20), Procrustes rotation yielded higher interfactor correlations than Oblimin rotation, although the differences in estimated correlations were minor (.015, .007, and .011 for CFA, and .011, .015, and .016 for PCA).

Finally, Table A7 of the supplementary material shows the simulation results when using MLFA instead of PAF as a model fitting procedure for CFA. MLFA generated a considerably higher percentage of improper solutions than PAF, reaching 86%–89% for three overextraction cases (Cases 17, 19, and 20; see Table A2). For cases in which PAF yielded higher *CC*s than PCA, MLFA tended to yield even higher *CC*s (e.g., Case 14: .835, .868, and .805 for PAF, MLFA, and PCA, respectively), whereas in cases in which PAF yielded lower *CC*s than PCA, MLFA yielded even lower *CC*s (e.g., Case 13: .842, .812, and .875 for PAF, MLFA, and PCA, respectively). In other words, the differences between CFA and PCA loadings were magnified when using MLFA instead of PAF as method of analysis. However, the overall pattern of differences between CFA and PCA was very similar for PAF and MLFA, as can be seen by the mean *Mnz* across all cases (.578, .580, and .639 for PAF, MLFA, and PCA, respectively) and the mean *CC* across all cases (.881, .883, and .876 for PAF, MLFA, and PCA, respectively).

4. Discussion

This simulation study confirmed earlier research by showing that PCA loadings are higher than CFA loadings, especially when the number of variables per factor is small. PCA has a

tendency to yield homogeneous loadings when factor loadings are unequal within factors. Furthermore, PCA loadings correlate relatively weakly with the true factor loadings for under- and overextraction. Our results extend the findings of Widaman (1993, 2007) to conditions with sampling error and under/overextraction. We also showed that PCA yields slightly higher *CCs* with the true factor loadings than CFA, for cross-loadings, model error, and factors of unequal strength.

Our simulation study showed that sampling error does not significantly influence the differences between CFA and PCA loadings, as the pattern of differences between CFA and PCA loadings was consistent for $N = 50$, $N = 200$, and $N = 50{,}000$. Specifically, PCA always yielded higher mean loadings than CFA and almost always lower standard deviations of nonzero loadings and higher standard deviations of zero loadings than CFA, regardless of the sample size. Our results showed that when sampling error was present, PCA occasionally provided a better approximation of the true factor loadings (in terms of *Mnz* and/or *CC*) than CFA did itself. For correlated factors in an otherwise clean factor pattern (Case 2), PCA yielded a higher *CC* with the true loadings, *and* mean loadings closer to the true factor loadings than CFA. This was true under both Oblimin and Procrustes rotation, indicating that the low loading estimates by CFA were not due to the rotation procedure. This superior performance of PCA comes artifactually in a sense, because sampling error results in lower loading estimates; the "inflated" loadings of PCA are a virtue in this circumstance.

PCA nonzero loadings were generally more stable than the corresponding CFA loadings. This result is in line with Ogasawara (2003), who derived mathematical expressions of the standard errors for factor and component loadings and showed that for equal numbers of variables per factor and equal factor loadings, the standard errors of component loadings are lower than the standard errors of factor loadings. However, our study also showed that the zero loading variables of PCA were less stable than those of CFA. Moreover, both nonzero and zero loading estimates of PCA were less stable than the corresponding loading estimates of CFA when departing from sphericity, that is, when loadings within factors were unequal.

For both CFA and PCA, underextraction appeared to be a more severe distortion than overextraction, in line with previous research (Fava and Velicer, 1996). Note that in all cases of underextraction (Cases 14, 15, and 18) we chose to make the omitted factors slightly weaker than the retained factors. When components are of equal strength (i.e., eigenvalues are equal), the principal axis cannot be uniquely determined in the population. This means that even the smallest degree of sampling error can completely change the eigenvector of the sample correlation matrix. CFA, on the other hand, iterates towards communalities that provided a good fit with the common factor model.

Because PCA loadings were higher than CFA loadings, correlations between components were lower than correlations between factors. For matrices in which each manifest variable loaded on one factor only, all true loadings on a factor were equal, and there was no sampling error, the predicted correlations between oblique components were in agreement with Widaman (1993). Both CFA and PCA yielded substantially lower correlations between factors/components in the case of sampling error than at the population level.

Our simulations showed that CFA is susceptible to improper solutions. Improper solutions with principal axis factoring (PAF) occurred particularly for overextraction and for a small number of variables per factor. Even for a sample size as large as 200, about 20% of the solutions were improper when overextracting two factors, or when having only two salient loadings on a factor. For maximum likelihood factor analysis (MLFA), the rate of improper solutions was even higher, being more than 50% in 6 of the 20 cases. If an improper solution occurred for CFA, the results of both methods were discarded, thereby

giving no penalty to CFA for producing an invalid outcome. PCA produced meaningful solutions when CFA failed to provide a proper solution (see Table 4). Some researchers have suggested that an improper solution has diagnostic value, indicating that the model attempted to be fit to the data is misspecified or that factor model assumptions are violated by the data, a problem that PCA will "bypass" (McArdle, 1990, p. 85). It is possible to prevent improper solutions by keeping the number of iterations of the principal axis factoring procedure small (cf. Snook and Gorsuch, 1989; Gorsuch, 2003), a pragmatic sanctioning of the common factor model.

Because exploratory CFA and PCA are typically applied to correlation matrices, the present simulations focused on correlation matrices only. MacCallum and Tucker (1991) explained that fitting the factor model to a correlation rather than a covariance matrix introduces additional sampling error, because the loadings are rescaled by rows as a function of the sample standard deviations of the manifest variables. Schneeweiss (1997) pointed out that the standardization of variables results in unique variances becoming more equal. Hence, differences between CFA and CFA loadings are smaller when the methods are applied on a correlation matrix instead of a covariance matrix. Further research needs to be conducted into the effect of the correlation versus covariance matrices in exploratory data analysis.

This simulation study cannot answer the infamous debate about whether CFA or PCA is the most appropriate procedure. The question about which method to use hinges on issues of generality, factor indeterminacy, and the theoretical status of latent variables (Borsboom et al., 2003; Steiger, 1994). An infinite number of factor scores can explain a given pattern of correlations—an indeterminacy issue discussed by many before (see a Special Section in *Multivariate Behavioral Research, 31 (4)*, "Indeterminacy," 1996). That is, although the loadings and the unique variances are uniquely determined from a given correlation matrix (yet not always, due to a small number of variables per factor or Heywood cases), the factor scores are not uniquely defined. Schönemann and Steiger (1978) argued that the factor model has "absurd properties" (p. 290), as the factor scores can be chosen such that they predict any criterion. Steiger (1994) pointed out that "going beyond the test space" may be merely "a theoretical advantage" of factor analysis, because "the vector . . . that takes them there is arbitrary. . . . the only determinate part of a common factor is, indeed, a component" (p. 213). Velicer and co-authors recommended "to do no analysis" (Velicer and Jackson, 1990b, p. 101) when CFA and PCA loadings differ, or alternatively, to improve the quality of the dataset by increasing the number of variables per factor or by including variables with high loadings. However, this recommendation exactly leads to a convergence of CFA and PCA loadings (cf. Eq. (10)). Steiger (1996) stated: "It seems high-reliability linear composites are what the factor analytic community has been looking for all along" (pp. 549–550). Researchers sometimes seem to let a factor analysis procedure defining factors for them. However, definitions should come from substantive evidence rather than from a correlation matrix alone. Recovering factors should be the result of both an appropriate selection of variables and of a data analytical technique that allows factors to have "some reality beyond the mathematical manipulations that derived them from a correlation matrix" (Jensen and Weng, 1994, p. 254, and see Johnson et al., 2004 for an interesting empirical study on the general intelligence factor).

Summarizing, the present work investigated differences between CFA and PCA loadings for a wide variety of conditions with and without sampling error. Results showed that CFA loadings were invariant to the number of variables in the dataset and that nonzero PCA loadings were generally higher than the true factor loadings, especially when the number of variables was small. Compared to CFA loadings, PCA loadings correlated weakly with the

true factor loadings for underextraction and overextraction. Heterogeneous factor loadings within correlated factors were homogenized by PCA. We found that PCA provided slightly higher *CCs* with the true factor loadings than CFA, for cross-loadings, model error, and factors of unequal strength. CFA occasionally suffered from a high rate of improper solutions. Sampling error led to a lowering of loadings and interfactor correlations for both CFA and PCA. The pattern of differences between CFA and PCA was consistent across sample sizes, levels of true factor loadings, principal axis factoring versus maximum likelihood factor analysis, and blind versus target rotation.

So, should a practitioner use CFA or PCA? Clearly, as pointed out by Widaman (2007), "the final word on comparisons between CFA and PCA has not yet been written" (p. 201). The present simulation study showed that PCA loadings are sometimes closer approximations of the true factor loadings than the loadings produced by CFA itself. However, this outperformance seems to be insufficient reason for justifying PCA as a substitute for CFA, because it is 'artifactually' caused by sampling error, is relatively rare (PCA outperformed CFA on *both Mnz* and *CC* only in Case 2), and probably of minor practical significance. We wish to reiterate that CFA and PCA are not competing techniques, as both methods facilitate a different purpose (explaining a correlation matrix vs. identifying the major sources of variation in data, respectively). The researcher's purpose should be the deciding factor in the choice of method. If a researcher has reasons to apply PCA, such as data reduction to a smaller dimensionality, then PCA should be used, and he should then not pretend that components are common factors. Practitioners seem to use CFA and PCA almost interchangeably, leading to heterogeneity in the scientific literature. The results of the present simulation study may contribute to measuring this heterogeneity and to increased awareness of the numerical consequences of using PCA or CFA.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website at http://dx.doi.org/10.1080/03610918.2013.862274

References

Anderson, T. W., Rubin, H. (1956). Statistical inference in factor analysis. In: Neyman J., ed., *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5, pp. 111–150.

Beauducel, A. (2001). On the generalizability of factors: The influence of changing contexts of variables on different methods of factor extraction. *Methods of Psychological Research Online* 6:69–96.

Bentler, P. M., Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research* 25:67–74.

Bernaards, C. A., Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement* 65:676–696.

Bookstein, F. L. (1990). Least squares and latent variables. *Multivariate Behavioral Research* 25:75–80.

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

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

Costello, A. B., Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation* 10.

De Winter, J. C. F., Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics* 39:695–710.

De Winter, J. C. F., Dodou, D., Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research* 44:147–181.

DiStefano, C., Zhu, M., Mîndrilã, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation* 14.

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research* 18:197–218.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* 4:272–299.

Fava, J. L., Velicer, W. F. (1992a). An empirical comparison of factor, image, component, and scale scores. *Multivariate Behavioral Research* 27:301–322.

Fava, J. L., Velicer, W. F. (1992b). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research* 27:387–415.

Fava, J. L., Velicer, W. F. (1996). The effects of underextraction in factor and component analyses. *Educational and Psychological Measurement* 56:907–929.

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

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

Glass, G. V., Taylor, P. A. (1966). Factor analytic methodology. *Review of Educational Research* 36:566–587.

Goldberg, L. R. (1990). An alternative "description of personality": The Big Five factor structure. *Personality Processes and Individual Differences* 59:1216–1229.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research* 25:33–39.

Gorsuch, R. L. (2003). Factor analysis. In: Schinka, J. A., Velicer, F. eds., *Handbook of Psychology: Vol. 2. Research Methods in Psychology.* Hoboken, NJ: John Wiley, pp. 143–164.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53:325–338.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods* 6:430–450.

Henson, R. K., Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement* 66:393–416.

Hong, S. (1999). Generating correlation matrices with model error for simulation studies in factor analysis: A combination of the Tucker-Koopman-Linn model and Wijsman's algorithm. *Behavior Research Methods, Instruments, & Computers* 31:727–730.

Indeterminacy. (1996). [Special commentary section]. *Multivariate Behavioral Research* 31:517–689.

Jensen, A. R. (1983). The definition of intelligence and factor-score indeterminacy. *The Behavioral and Brain Sciences* 6:313–315.

Jensen, A. R., Weng, L.-J. (1994). What is a good *g*? *Intelligence* 18:231–258.

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence* 32:95–107.

Jolliffe, I. T. (2002). *Principal Component Analysis* 2nd ed. New York: Springer.

Kaiser, H. F., Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika* 27:179–182.

Kaplunovsky, A. S. (2005). Factor analysis in environmental studies. *HAIT Journal of Science and Engineering B* 2:54–94.

Lawrence, F. R., Hancock, G. R. (1999). Conditions affecting integrity of a factor solution under varying degrees of overextraction. *Educational and Psychological Measurement* 59:549–579.

Loehlin, J. C. (1990). Component analysis versus common factor analysis: A case of disputed authorship. *Multivariate Behavioral Research* 25:29–31.

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.

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

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

Maraun, M. D. (1996). Metaphor taken as math: indeterminancy in the factor analysis model. *Multivariate Behavioral Research* 31:517–538.

McArdle, J. J. (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research* 25:81–87.

McCloy, C. H., Metheny, A., Knott, V. (1938). A comparison of the Thurnstone method of multiple factors with the Hotelling method of principal components. *Psychometrika* 3:61–67.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research* 22:267–305.

Mulaik, S. A. (1990). Blurring the distinction between component analysis and common factor analysis. *Multivariate Behavioral Research* 25:53–59.

Mulaik, S. A. (2010). *Foundations of Factor Analysis.* 2nd ed. Boca Raton, FL: Chapman and Hall/CRC.

Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika* 65:167–185.

Ogasawara, H. (2003). Oblique factors and components with independent clusters. *Psychometrika* 68:299–321.

Park, H. S., Dailey, R., Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research* 28:562–577.

Parmet, Y., Schechtman, E., Sherman, M. (2010). Factor analysis revisited—How many factors are there? *Communications in Statistics—Simulation and Computation* 39:1893–1908.

Paschou, P., Ziv, E., Burchard, E. G., Choudhry, S., Rodriguez-Cintron, W., Mahoney, M. W., Drineas, P. (2007). PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genetics* 3:1672–1686.

Peterson, R. A. (2000). A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters* 11:261–275.

Preacher, K. J., Zhang, G., Kim, C., Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research* 48:28–56.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika* 20:93–111.

Rao, C. R. (1996). *Principal Component and Factor Analyses (No. TR-96-04).* Pennsylvania State University: Center for Multivariate Analysis. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA313617

Reich, D., Price, A. L., Patterson, N. (2008). Principal component analysis of genetic data. *Nature Genetics* 40:491–492.

Rozeboom, W. W. (1990). Whatever happened to broad perspective? *Multivariate Behavioral Research* 25:61–65.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin*. *Personality and Social Psychology Bulletin* 28:1629–1646.

Sato, M. (1990). Some remarks on principal component analysis as a substitute for factor analysis in monofactor cases. *Journal of the Japan Statistical Society* 20:23–31.

Schmidt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment* 29:304–321.

Schönemann, P. H. (1990). Facts, fictions, and common sense about factors and components. *Multivariate Behavioral Research* 25:47–51.

Schönemann, P. H., Steiger, J. H. (1978). On the validity of indeterminate factor scores. *Bulletin of the Psychonomic Society* 12:287–290.

Schneeweiss, H. (1997). Factors and principal components in the near spherical case. *Multivariate Behavioral Research* 32:375–401.

Schneeweiss, H., Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis* 55:105–124.

Snook, S. C., Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin* 106:148–154.

Steiger, J. H. (1990). Some additional thoughts on components, factors, and factor indeterminacy. *Multivariate Behavioral Research* 25:41–45.

Steiger, J. H. (1994). Factor analysis in the 1980's and the 1990's: Some old debates and some new developments. In: Borg, I. Mohler, P. Ph. eds., *Trends and Perspectives in Empirical Social Research.* Berlin: Walter de Gruyter, pp. 201–224.

Steiger, J. H. (1996). Dispelling some myths about factor indeterminancy. *Multivariate Behavioral Research* 31:539–550.

Steiger, J. H. (2004). Paul Meehl and the evolution of statistical methods in psychology. *Applied & Preventive Psychology* 11:69–72.

Trendafilov, N. T., Unkel, S., Krzanowski, W. (2013). Exploratory factor and principal component analyses: some new aspects. *Statistics and Computing* 23:209–220.

Tucker, L. R. (1951). *A Method for Synthesis of Factor Analysis Studies (Personnel Research Section Report No. 984)*. Washington, DC: Department of the Army.

Tucker, L. R, Koopman, R. F., Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika* 34:421–459.

Velicer, W. F. (1974). A comparison of the stability of factor analysis, principal component analysis, and rescaled image analysis. *Educational and Psychological Measurement* 34:563–572.

Velicer, W. F. (1976a). Determining the number of components from the matrix of partial correlations. *Psychometrika* 41:321–327.

Velicer, W. F. (1976b). The relation between factor score estimates, image scores, and principal component scores. *Educational and Psychological Measurement* 36:149–159.

Velicer, W. F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterns. *The Journal of Multivariate Behavioral Research* 12:3–22.

Velicer, W. F., Fava, J. L. (1987). An evaluation of the effects of variable sampling on component, image, and factor analysis. *Multivariate Behavioral Research* 22:193–209.

Velicer, W. F., Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods* 3:231–251.

Velicer, W. F., Jackson, D. N. (1990a). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research* 25:1–28.

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

Velicer, W. F., Peacock, A. C., Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research* 17:371–388.

Widaman, K. F. (1990). Bias in pattern loadings represented by common factor analysis and component analysis. *Multivariate Behavioral Research* 25:89–95.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research* 28:263–311.

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In: Cudeck, R. MacCallum, R. C. eds., *Factor Analysis at 100: Historical Developments and Future Directions,* Mahwah, NJ: Lawrence Erlbaum Associates, pp. 177–203.