

Sample Size in Factor Analysis: The Role of Model Error

Robert C. MacCallum
Ohio State University

Keith F. Widaman
University of California, Davis

Kristopher J. Preacher
Ohio State University

Sehee Hong
University of California, Santa Barbara

This article examines effects of sample size and other design features on correspondence between factors obtained from analysis of sample data and those present in the population from which the samples were drawn. We extend earlier work on this question by examining these phenomena in the situation in which the common factor model does not hold exactly in the population. We present a theoretical framework for representing such lack of fit and examine its implications in the population and sample. Based on this approach we hypothesize that lack of fit of the model in the population will not, on the average, influence recovery of population factors in analysis of sample data, regardless of degree of model error and regardless of sample size. Rather, such recovery will be affected only by phenomena related to sampling error which have been studied previously. These hypotheses are investigated and verified in two sampling studies, one using artificial data and one using empirical data.

This article extends previous studies of the question of the level of sample size (N) in factor analysis studies that is necessary to achieve accurate recovery of major common factors present in the population. This question has been examined in a variety of studies over many years. Simulation studies (e.g., Browne, 1968; Pennell, 1968; Velicer, Peacock, & Jackson, 1982) have investigated effects of sample size on recovery of population factors, as well as interactions between sample size and other factors such as loading size, number of variables, and number of factors. Some sampling studies using empirical data (e.g., Arrindell & van der Ende, 1985; Barrett & Kline, 1981) have examined congruence between factors

Correspondence regarding this article should be directed to Robert MacCallum, Department of Psychology, 1885 Neil Avenue, Ohio State University, Columbus, OH 43221. E-mail: maccallum.1@osu.edu.

obtained from an extremely large sample and those obtained from subsamples of various sizes. In addition, investigators have proposed a range of rules of thumb for minimum N , specifying minimum N as a function of number of variables, number of parameters being estimated, or as some absolute level (e.g., Cattell, 1978; Comrey & Lee, 1992; Gorsuch, 1983; Guilford, 1954).

Much of this literature has been reviewed in recent articles by MacCallum, Widaman, Zhang, and Hong (1999) and Velicer and Fava (1998). Both studies showed that rules of thumb are not valid and that the minimum level of N is dependent on other aspects of design. Level of communality has an especially strong interaction with N such that when communalities are high, good recovery of population factors can be achieved with relatively small samples. However, when communalities are low, recovery of population factors is difficult to achieve unless N is extremely large. MacCallum et al. also showed how these aspects of design interact with level of overdetermination of factors, meaning the degree to which each common factor is well defined by a set of indicators. The main effects of N and communality level on recovery of population factors are more dramatic when factors are less well determined.

Based on earlier work by MacCallum and Tucker (1991) on sources of error in factor analysis, MacCallum et al. (1999) provided a formal theoretical framework to account for these effects. These phenomena were verified in a Monte Carlo study, with results consistent with findings by Velicer and Fava (1998). A limitation of that theoretical framework and of the Monte Carlo studies conducted by MacCallum et al. and by Velicer and Fava is that all of that work was based on the ideal case of exact fit of the common factor model in the population. This same limitation applies to all previous Monte Carlo studies of the sample size question in factor analysis of which we are aware. This limitation is of concern because it represents a condition that will almost certainly never hold in practice. A widely accepted perspective regarding mathematical models of psychological phenomena is that such models will never hold exactly in practice. In the context of factor analysis and structural equation modeling, for example, this perspective has been discussed in detail by Cudeck and Henly (1991) and MacCallum and Tucker (1991), among others. Considering here the common factor model, there are many influences that will cause the model not to hold in practice. These include, for instance, the influences of large numbers of minor factors, nonlinear influences of factors on indicators, and violation of the assumption that factor loadings are the same for every individual. For present purposes, we lump these various influences together and refer to them in the aggregate as "model error." This same aggregate

lack of fit has been called “discrepancy of approximation” by Cudeck and Henly in the context of covariance structure modeling. Because these and other phenomena are likely to be operating in any real-world setting, the best we can hope for in empirical studies is to obtain a parsimonious model that fits the data well and yields interpretable parameter estimates. If such a finding is obtained, we must be fully aware that the resulting model and solution are not an exact or complete representation of phenomena operating in the real world, but only a parsimonious approximation.

Previous studies of the sample size question in factor analysis have not explicitly taken into account the existence or influence of model error. In particular, the mathematical framework presented by MacCallum et al. (1999) was based on an assumption that the model holds exactly in the population. Simulation studies conducted by MacCallum et al. and by Velicer and Fava (1998) used simulated population correlation matrices whose structure was exactly consistent with the common factor model using a specified (small) number of factors. In the present study we address this limitation. We first extend the theoretical framework used by MacCallum et al. so as to represent model error explicitly, and we use this extended version to predict effects of sample size and other aspects of design in the presence of such error. We then investigate these predictions in two sampling studies, one using artificial data containing model error and one using empirical data. Given the prevalence of model error in applied work using factor analysis, the outcomes of the present study are more relevant than are previous results to the sample size question in empirical applications of factor analysis.

Theoretical Framework

We first review key aspects of the mathematical approach used by MacCallum et al. (1999), which was extracted from more general earlier work by MacCallum and Tucker (1991). For a column vector \mathbf{y} of scores on p measured variables for a random individual, let the common factor model be represented as follows:

$$(1) \quad \mathbf{y} = \mathbf{\Lambda}\mathbf{x}_c + \mathbf{\Theta}\mathbf{x}_u$$

where $\mathbf{\Lambda}$ is a $p \times r$ matrix of loadings for p measured variables on r common factors, $\mathbf{\Theta}$ is a $p \times p$ diagonal matrix of unique factor loadings, \mathbf{x}_c is a vector of scores on the r common factors for the random individual, and \mathbf{x}_u is a vector of scores on the p unique factors for the random individual. All variables are assumed to have zero means. Defining unique factors as

uncorrelated with each other and with common factors in the population, we can derive the following covariance structure from Equation 1:

$$(2) \quad \Sigma_{yy} = \Lambda\Phi\Lambda' + \Theta^2$$

where Σ_{yy} is the $p \times p$ population covariance matrix for the measured variables and Φ is the $r \times r$ population correlation matrix for the common factors (assuming factors are standardized in the population). This is the standard version of the common factor model for a population covariance matrix. Following similar algebraic procedures, we could derive a structure for a *sample* covariance matrix, C_{yy} . However, in such a derivation we can not assume that unique factors are uncorrelated with each other or with common factors *in the sample*. Although such correlations will be zero in the population, they will not generally be zero in a sample due to random sampling variability. Thus, the structure of C_{yy} would be given by

$$(3) \quad C_{yy} = \Lambda C_{cc} \Lambda' + \Lambda C_{cu} \Theta + \Theta C_{uc} \Lambda' + \Theta C_{uu} \Theta$$

Matrices C_{cc} , C_{cu} , C_{uc} , and C_{uu} are sample covariance matrices among common and unique factors as indicated by subscripts c and u , respectively. If Equation 1 holds for all individuals in the population, then Equation 3 holds for any sample from that population. The most important aspect of Equation 3 involves the presence of nonzero sample correlations among unique factors in C_{uu} , and of unique with common factors in C_{cu} and C_{uc} . This phenomenon is the reason that the model in Equation 2 will not hold in a sample, even if Equation 1 holds for all individuals. Thus, the existence of nonzero sample correlations among unique factors, and of unique with common factors, represents a primary aspect of sampling error in the common factor model. If sample covariances among common and unique factors matched corresponding population values, the model in Equation 2 would apply to a sample covariance matrix as well as to a population covariance matrix. This circumstance suggests an alternative representation of Equation 3 as follows:

$$(4) \quad C_{yy} = \Lambda\Phi\Lambda' + \Theta^2 + \Delta_{SE}$$

where Δ_{SE} represents lack of fit of the model in Equation 3 arising from this primary source of sampling error (hence the subscript SE). If sample covariances involving common and unique factors were to match exactly their population values (that is, $C_{cc} = \Phi$; C_{cu} and C_{uc} are zero; and $C_{uu} = \mathbf{I}$), then Δ_{SE} would be null. Clearly, the most critical phenomenon here is the

existence in the sample of nonzero covariances of unique factors with each other and with common factors, rendering \mathbf{C}_{cu} and \mathbf{C}_{uc} nonzero and \mathbf{C}_{uu} non-diagonal.

MacCallum et al. (1999) used this perspective to examine the impact of sample size on factor analysis solutions. As N increases, sample variances and covariances among unique and common factors will approach population values, thus reducing the impact of Δ_{SE} and causing sample factor analysis solutions to be more similar to population solutions. They also noted, based on Equation 3, that the magnitude of unique factor weights in Θ plays a critical role. As those weights decrease (or, equivalently, as communalities increase), the contents of \mathbf{C}_{cu} , \mathbf{C}_{uc} , and \mathbf{C}_{uu} become less relevant, meaning that sample size will have less impact on factor analysis solutions when communalities are high. MacCallum et al. also discussed the issue of overdetermination of factors, meaning the extent to which each factor is represented by a distinct set of indicators. They noted that recovery of population factors will deteriorate when factors are weakly overdetermined, but that this phenomenon would also be attenuated when communalities are high. MacCallum et al. verified these predicted effects in a Monte Carlo study using simulated population data drawn from the classic study of factor analysis methods by Tucker, Koopman, and Linn (1969). Results verified that, when communalities are high, sample factor solutions correspond closely to population solutions even when N is small and factors are weakly overdetermined. When communalities are not high, sample size and level of overdetermination have much stronger effects on quality of sample solutions. Similar phenomena were reported by Velicer and Fava (1998).

As noted earlier, the theoretical framework used by MacCallum et al. (1999) and summarized above, as well as the Monte Carlo studies of MacCallum et al. and of Velicer and Fava (1998), are based on the assumption that the common factor model holds exactly in the population. We next consider extending that work to the case where lack of fit, or model error, is present in the population so that we can determine whether effects of sample size, communality level, and overdetermination are altered in that more realistic context.

Following the approach used by MacCallum and Tucker (1991), we begin by modifying the basic model in Equation 1 as follows:

$$(5) \quad \mathbf{y} = \Lambda \mathbf{x}_c + \Theta \mathbf{x}_u + \mathbf{y}_d$$

where \mathbf{y}_d is a vector containing that part of the measured variables in \mathbf{y} that is not accounted for by the common factor model; that is, \mathbf{y}_d reflects discrepancies between the model and the measured variables. Note that \mathbf{y}_d

does not contain portions due to error of measurement or specific variances; those aspects of \mathbf{y} are accounted for by the unique factor portion of the model in Equation 5. Rather, \mathbf{y}_d arises from phenomena such as those mentioned above; for example, influences of minor common factors and nonlinear influences of common factors on measured variables. That portion of the variables in \mathbf{y} that is explained by the model will be designated \mathbf{y}_m and represented as follows:

$$(6) \quad \mathbf{y}_m = \Lambda \mathbf{x}_c + \Theta \mathbf{x}_u$$

The variables in \mathbf{y}_m were referred to as “modeled variables” by MacCallum and Tucker (1991). Substitution of Equation 6 into 5 yields

$$(7) \quad \mathbf{y} = \mathbf{y}_m + \mathbf{y}_d$$

representing the measured variables in \mathbf{y} as a sum of two parts, that part accounted for by the common factor model and that part not accounted for.

It is instructive to consider the structure of the population covariance matrix of the measured variables in the presence of model error. From Equation 7, this covariance matrix Σ_{yy} could be represented as a sum of four matrices as follows:

$$(8) \quad \Sigma_{yy} = \Sigma_{y_m y_m} + \Sigma_{y_m y_d} + \Sigma_{y_d y_m} + \Sigma_{y_d y_d}$$

The first of these four component matrices is the covariance matrix for the modeled variables, which would be exactly fit by the common factor model:

$$(9) \quad \Sigma_{y_m y_m} = \Lambda \Phi \Lambda' + \Theta^2$$

The remaining three component matrices arise from lack of fit of the model in the population. They are not present in this formulation if the model holds exactly in the population. These three components can be aggregated into a single matrix, $\Delta_{ME(P)}$, representing model error in the population covariance structure:

$$(10) \quad \Delta_{ME(P)} = \Sigma_{y_m y_d} + \Sigma_{y_d y_m} + \Sigma_{y_d y_d}$$

Substituting from Equations 9 and 10 into Equation 8 yields a representation of the factorial structure of a population covariance matrix in the presence of model error:

$$(11) \quad \Sigma_{yy} = \Lambda\Phi\Lambda' + \Theta^2 + \Delta_{ME(P)}$$

In one sense this expression is intuitively obvious in that matrix $\Delta_{ME(P)}$ can be viewed as representing lack of fit of the model specified earlier in Equation 2. However, the developments in Equations 6 through 11 provide a more explicit representation of the nature of this lack of fit. In addition, as shall be seen, this approach aids in providing a basis for understanding effects of sample size in factor analysis of sample data.

Let us consider next the structure of a sample covariance matrix in the presence of model error as defined in Equation 7. Based on that equation, we can represent a sample covariance matrix, C_{yy} , as a sum of four component matrices:

$$(12) \quad C_{yy} = C_{y_m y_m} + C_{y_m y_d} + C_{y_d y_m} + C_{y_d y_d}$$

Note that matrix $C_{y_m y_m}$ would have the structure represented in Equations 3 and 4; that is,

$$(13) \quad C_{y_m y_m} = \Lambda C_{cc} \Lambda' + \Lambda C_{cu} \Theta + \Theta C_{uc} \Lambda' + \Theta C_{uu} \Theta \\ = \Lambda\Phi\Lambda' + \Theta^2 + \Delta_{SE}$$

Recall that Δ_{SE} represents lack of fit due to sampling error, arising primarily from nonzero sample covariances of unique factors with each other and with common factors. Considering further Equation 12 we can identify another source of lack of fit of the model in a sample. The last three component covariance matrices in Equation 12 arise from the presence of model error in the sample; these matrices would not be present if the model in Equation 1 held for every individual. These three matrices can be considered in the aggregate as defining a matrix $\Delta_{ME(S)}$ representing lack of fit due to effects of model error in the sample:

$$(14) \quad \Delta_{ME(S)} = C_{y_m y_d} + C_{y_d y_m} + C_{y_d y_d}$$

Substitution from Equations 13 and 14 into Equation 12 yields the following expression:

$$(15) \quad C_{yy} = \Lambda\Phi\Lambda' + \Theta^2 + \Delta_{SE} + \Delta_{ME(S)}$$

This expression represents the structure of a sample covariance matrix and includes terms representing lack of fit of the common factor model due to two sources of error: sampling error and model error. Of course in practice these

two sources of error act in concert to produce overall lack of fit of the model in a sample.

Based on this framework for representing sources of error in the common factor model in both the population and sample, we now examine the issue of correspondence between factors present in the population and those obtained from sample data, focusing on the role of sample size. Consider first Equation 11 where model error in the population is explicitly represented. According to this equation, a model could be fit to Σ_{yy} and the solution would provide population values of parameters in Λ , Φ , and Θ^2 . It is interesting to note that the particular set of parameter values obtained depends on the method used to fit the model to Σ_{yy} . That is, different methods (e.g., maximum likelihood, ordinary least squares) would yield different residual matrices $\Delta_{ME(P)}$ and different values of parameters in Λ , Φ , and Θ^2 . From this perspective, there is no single true set of parameter values when the model does not hold exactly in the population. Rather, *parameter values depend on the way in which lack of fit is defined*. This phenomenon was discussed and demonstrated by MacCallum and Tucker (1991). For present purposes, suppose a particular fitting method is used, thus yielding a specific set of parameter values. We focus especially on factor loadings in Λ because we are interested primarily in recovery of factors present in the population.

Next consider Equation 15, representing the structure of the sample covariance matrix C_{yy} . If the model is fit to C_{yy} by the same method used to fit the model to Σ_{yy} , we obtain factor loading estimates, $\hat{\Lambda}$, and we are interested in the correspondence between $\hat{\Lambda}$ and Λ . A comparison of Equations 11 and 15 provides a basis for understanding what would cause these solutions to differ. Clearly, any difference between these solutions arises from the role of the two discrepancy terms: $\Delta_{ME(P)}$ in the population, and $[\Delta_{SE} + \Delta_{ME(S)}]$ in the sample. If these two discrepancy terms were identical, then $\hat{\Lambda}$ would match Λ , implying exact correspondence of sample and population factors. However, to the extent that the two discrepancy terms differ, the sample solution will differ from the population solution, resulting in poorer correspondence between factors found in the sample and those present in the population. Thus, we must examine phenomena that affect the similarity of the two discrepancy terms. Most apparent is the presence of the sampling error term Δ_{SE} in the residual portion of the model for C_{yy} . Obviously, sampling error affects a solution obtained from C_{yy} , but not one obtained from Σ_{yy} . MacCallum et al. (1999) examined these influences in detail, describing and demonstrating effects of N , communality level, and degree of overdetermination of factors as mentioned above. The present developments

indicate that these phenomena should operate in the same manner when model error is present as when it is absent (as examined in previous studies), as indicated by the difference between Equations 11 and 15.

The second major difference between the lack-of-fit terms in Equations 11 and 15 is the distinction between model error in the population [$\Delta_{ME(P)}$] and model error in the sample [$\Delta_{ME(S)}$]. These terms were defined in Equations 10 and 14, respectively. To the extent that these terms are similar, population factors will be recovered more accurately in analysis of sample data. If these terms were to match exactly, then differences between sample and population factors would be due only to sampling error. What would make these terms be similar? First, they will be more similar as model error becomes small in magnitude. In such a situation, $\Delta_{ME(P)}$ and $\Delta_{ME(S)}$ would both be very small and would have a negligible impact on results. Second, these matrices will be more similar as N becomes large. That is, $\Delta_{ME(S)}$ will approach $\Delta_{ME(P)}$ as N increases. However, a more critical point is that, regardless of N , we would expect that, over repeated sampling, the average contribution of model error in the sample will approximately equal that in the population. This means that, on average, *model error will not influence the difference between factors obtained from a sample and those present in the population.* In turn, this means that, on average, *such differences will be due to influences of sampling error only.*

To summarize, over repeated sampling, we hypothesize that mean recovery of population factors will not be influenced by the presence of model error. Rather, on average, such recovery will be influenced only by sampling error, with effects corresponding to those demonstrated by MacCallum et al. (1999) in their study of recovery of factor solutions in the absence of model error. In the remainder of this article, these hypotheses are investigated in two sampling studies, the first using simulated data and the second using a large set of empirical data.

Monte Carlo Study

MacCallum et al. (1999) conducted a Monte Carlo study to investigate effects of sample size, communality, and overdetermination on recovery of population factors under conditions in which the common factor model holds in the population. We followed similar procedures to investigate these effects under conditions in which the model does not hold exactly in the population. We then compared our results with their results to verify the predicted effects described in the previous section.

Method

In their study, MacCallum et al. (1999) used selected population correlation matrices generated in the Monte Carlo study of factor analytic methods by Tucker et al. (1969). Those selected matrices were generated from the common factor model with no model error. Tucker et al. also constructed population correlation matrices for which the model did not hold exactly. This was accomplished by simulating effects of large numbers of minor common factors. For full details about the procedure used to generate these population correlation matrices, refer to Tucker et al. (1969).

In the present project we generated 18 population correlation matrices using methods described by Tucker et al. (1969). As in the MacCallum et al. (1999) study, the matrices varied in number of factors and level of communality, but the number of measured variables (p) was held constant at 20. Number of factors was either three (highly overdetermined) or seven (relatively weakly overdetermined). For a given number of variables and factors, the Tucker et al. simulation procedure generates a population factor loading matrix that exhibits good, but not perfect, simple structure. Like loading matrices typically encountered in empirical studies, these simulated matrices exhibit variation in importance of major factors as well as numerous small to moderate secondary loadings. Following Tucker et al., we varied the contribution of these major common factors in accounting for the variance of each measured variable. Tucker et al. referred to this quantity as b_1^2 and assigned a value of b_1^2 to each measured variable. We used the same values of b_1^2 as did Tucker et al. In generating a given matrix, the level of b_1^2 for each measured variable was specified as high (.6 to .8), wide (.2 to .8), or low (.2 to .4). Given the presence of minor common factors, as described below, the level of b_1^2 will not be equivalent to level of communality, although communalities vary with b_1^2 . In the present study, population correlation matrices generated under high, wide, and low levels of b_1^2 yielded mean communalities of .70, .55, and .35, respectively. For simplicity of presentation, we will henceforth refer to our Monte Carlo design as including high, wide, and low levels of communality. The communality levels correspond inversely to levels of importance of unique factors. High communalities imply low unique variances and vice versa.

In addition to manipulating overdetermination and communality levels, we controlled the degree of model fit in the population. The Tucker et al. (1969) procedure for introducing lack of fit involved simulating effects of a large number (set at 180) of minor factors. Degree of misfit can be manipulated by varying the proportion of variance in each measured variable attributable to minor factors as well as a data generation parameter (called

ϵ by Tucker et al.) that determines the relative contribution of successive minor factors; larger values of ϵ cause the contribution of the minor factors to be more skewed in favor of the earlier factors in the sequence. In the present study, we manipulated ϵ and the proportion of variance due to minor factors so as to yield population correlation matrices exhibiting good, fair, and poor fit, respectively, to the common factor model. We used the root mean squared error of approximation (RMSEA, Steiger & Lind, 1980) as an index of model fit. Steiger (1989) and Browne and Cudeck (1993) suggest guidelines for the interpretation of RMSEA: values in the range of 0.00 to 0.05 indicate close fit, those between 0.05 and 0.08 indicate fair fit, and those between 0.08 and 0.10 indicate mediocre fit. RMSEA values above 0.10 indicate unacceptable fit. In order to represent the broad range of model misfit researchers are likely to encounter in practice, we generated population matrices with RMSEA values corresponding to the mid-point of each of the “acceptable” ranges (very good fit: RMSEA = 0.025; fair fit: RMSEA = 0.065; poor fit: RMSEA = 0.090). One population correlation matrix was selected to represent each condition defined by two levels of number of factors, three levels of communality, and three levels of model misfit, yielding 18 population correlation matrices.¹

Sample correlation matrices were generated from each of these 18 population correlation matrices using a procedure suggested by Wijsman (1959) and used in the MacCallum et al. (1999) study. Population distributions were defined as multivariate normal. Sample size was 60, 100, 200, or 400. These sample sizes were chosen for comparison with results obtained by MacCallum et al. (1999). We generated 100 samples for each condition defined by three levels of communality, two levels of overdetermination, three levels of model misfit, and four levels of sample size. Each sample correlation matrix was factor analyzed using maximum likelihood factor analysis, specifying the number of factors retained as equal to the known number of factors in the population (i.e., either three or seven). All factor analyses were conducted using the Comprehensive Exploratory Factor Analysis (CEFA) software of Tateneni, Mels, Cudeck, and Browne (1998).

In conducting these factor analyses it was necessary to determine how best to deal with solutions that yielded Heywood cases or that had not converged (after a limit of 250 iterations). MacCallum et al. (1999) conducted their entire study twice, once screening such samples and generating new samples until each cell was filled with samples yielding convergent and proper solutions, and again with no such screening.

¹ The 18 population matrices, along with corresponding population factor patterns and communalities, can be found at the website of Robert C. MacCallum (<http://quantrm2.psy.ohio-state.edu/maccallum>).

Screening was found to have virtually no effect on results. In the current study, we carried out analyses both with and without screening in two cells expected to yield large numbers of solutions with Heywood cases (low communalities, $p:r = 20:7$, $N = 60$, poor fit and fair fit). Again, results in terms of indexes described below were nearly identical whether or not samples were screened. Given this finding along with results described by MacCallum et al., we conducted our full-scale study without screening.

We followed procedures used by MacCallum et al. (1999) to compare solutions obtained from sample correlation matrices with those obtained from the corresponding population matrices. Maximum likelihood factor solutions for the population correlation matrices were obtained, and these population factors were rotated by direct quartimin rotation. Maximum likelihood solutions obtained from sample correlation matrices were rotated by oblique least-squares target rotation, using the corresponding rotated population factor loading matrix as the target.² To assess similarity between factors obtained in the population and in a sample, we first calculated the coefficient of congruence (ϕ) between each population factor and the corresponding sample factor. Following MacCallum et al. (1999), a summary index of recovery of population factors in each sample solution was defined as the mean coefficient of congruence across the r (three or seven) factors. This index is designated as K :

$$(16) \quad K = \frac{\sum_{k=1}^r \phi_k}{r}.$$

Higher values of K indicate more accurate recovery of the population factors in the sample solution. Tucker (personal communication, 1987) suggested guidelines for interpreting the value of K : .98 to 1.00 = *excellent*, .92 to .98 = *good*, .82 to .92 = *borderline*, .68 to .82 = *poor*, and below .68 = *terrible*.

To further assess the degree of factor recovery, we computed an additional index of the discrepancy between sample factor solutions and their corresponding population factor solutions: the *root mean squared deviation*

² All subsequent analyses were also conducted using sample solutions rotated by direct quartimin rather than target rotation. The general pattern of results was not affected by this difference in rotation, although some trends were slightly more noisy. Plots of results based on direct quartimin rotation are available at the website of Robert C. MacCallum (<http://quantrm2.psy.ohio-state.edu/maccallum>).

(designated g , as in Velicer & Fava, 1998). This index represents the degree of discrepancy between corresponding sample and population loading matrices:

$$(17) \quad g = \left\{ \frac{\text{Trace} \left[(\Lambda - \hat{\Lambda})' (\Lambda - \hat{\Lambda}) \right]}{pr} \right\}^{1/2} .$$

To investigate possible bias in estimates of factor loadings, we calculated the mean deviation between sample loadings and their corresponding population loadings. This index is designated as δ :

$$(18) \quad \delta = \frac{\sum_{i=1}^p \sum_{j=1}^r (\lambda_{ij} - \hat{\lambda}_{ij})}{pr}$$

In summary, 100 sample correlation matrices were generated under each of 72 conditions defined by (a) three levels of population communality (high, wide, low), (b) two levels of overdetermination of the factors ($p:r = 20:3, 20:7$), (c) three levels of model misspecification (RMSEA = .025, .065, .090), and (d) four levels of sample size ($N = 60, 100, 200, 400$). The resulting 7,200 sample matrices were each analyzed using maximum likelihood factor analysis with the known correct number of factors specified. Each of the resulting 7,200 solutions was rotated to the corresponding direct-quartimin population solution, using oblique target rotation. For each of the rotated sample solutions, measures of congruence between sample and population factors (K and g) and bias (δ) were obtained.

Results

We wished to investigate whether the conclusions of MacCallum et al. (1999) regarding influences on recovery of population factors still hold under conditions where model error is present. Indexes K , g , and δ were treated as dependent variables in four-way ANOVAs using sample size, level of communality, overdetermination, and population model fit as independent variables. Given the large number of replications in the full design (7,200), it

was anticipated that virtually all main effects and interactions would be statistically significant. Therefore, our interpretation of effects was based primarily on effect size as measured by $\hat{\omega}^2$. This measure provides an estimate of the proportion of variance accounted for in the population by each effect (Maxwell & Delaney, 1990).

Table 1 presents results of the four-way ANOVA³ for the congruence measure, *K*. Note that all main effects and nearly all interactions were statistically significant. The largest effect was the main effect of level of communality ($\hat{\omega}^2 = .42$), and substantial effect-size estimates also were obtained for overdetermination ($\hat{\omega}^2 = .23$) and for sample size ($\hat{\omega}^2 = .27$), just as in the MacCallum et al. (1999) Monte Carlo study. Effect sizes for all

Table 1
ANOVA Results for the Measure of Congruence (*K*) in Monte Carlo Study

Source	<i>df</i>	<i>SS</i>	<i>F</i>	<i>Prob.</i>	$\hat{\omega}^2$
Sample size (<i>N</i>)	3	806.25	12663.70	<.0001	.27
Communality (<i>h</i>)	2	1260.63	29700.90	<.0001	.42
Overdetermination (<i>d</i>)	1	695.08	32752.82	<.0001	.23
Model fit (<i>f</i>)	2	18.45	434.64	<.0001	.01
<i>d</i> × <i>f</i>	2	7.70	181.38	<.0001	.00
<i>h</i> × <i>f</i>	4	17.52	206.33	<.0001	.00
<i>h</i> × <i>d</i>	2	26.54	625.25	<.0001	.01
<i>N</i> × <i>f</i>	6	0.25	1.94	0.071	.00
<i>N</i> × <i>d</i>	3	7.69	120.72	<.0001	.00
<i>N</i> × <i>h</i>	6	3.64	28.59	<.0001	.00
<i>h</i> × <i>d</i> × <i>f</i>	4	34.35	404.63	<.0001	.01
<i>N</i> × <i>d</i> × <i>f</i>	6	0.09	0.72	0.636	.00
<i>N</i> × <i>h</i> × <i>d</i>	6	4.51	35.41	<.0001	.00
<i>N</i> × <i>h</i> × <i>f</i>	12	0.98	3.86	<.0001	.00
<i>N</i> × <i>h</i> × <i>d</i> × <i>f</i>	12	1.17	4.60	<.0001	.00
Error	7128	151.27			

Note. *Prob.* = probability; $\hat{\omega}^2$ = estimated proportion of variance accounted for in the population by each effect.

³ Because the distribution of ϕ becomes more skewed as ϕ approaches 1.0, values of ϕ were transformed into normal deviates by means of the Fisher *r* to *z* transformation before inclusion of the *K* index in ANOVAs.

interactions were small. Importantly, even though the main effect for model error was statistically significant, its effect-size was quite small ($\hat{\omega}^2 = .0061$).

Table 2 presents results of the four-way ANOVA for the root-mean-square deviation measure, g . As with K , all main effects and nearly all interactions were statistically significant. The largest effect was the main effect of sample size ($\hat{\omega}^2 = .41$), followed closely by that of communality level ($\hat{\omega}^2 = .33$). Overdetermination had a lower effect-size estimate ($\hat{\omega}^2 = .07$), and the effect-sizes for all other main effects and interactions were small to negligible. As with K , even though the main effect for model error was statistically significant, its effect-size was quite small ($\hat{\omega}^2 = .0021$).

Cell means for K are presented in Figure 1, and cell means for g are presented in Figure 2. The error bars around each point in all of the plots in Figures 1 and 2 represent intervals within which approximately 95% of the values of K (or g) lie for each cell mean (e.g., $\bar{K} \pm 1.96S_{\bar{K}}$). As hypothesized, all of the main effects and interactions reported in MacCallum et al. (1999)

Table 2
ANOVA Results for the Root Mean Squared Error (g) in Monte Carlo Study

Source	<i>df</i>	<i>SS</i>	<i>F</i>	<i>Prob.</i>	$\hat{\omega}^2$
Sample size (<i>N</i>)	3	4.12	7837.64	<.0001	.41
Communality (<i>h</i>)	2	3.36	9588.25	<.0001	.33
Overdetermination (<i>d</i>)	1	0.79	4531.38	<.0001	.08
Model fit (<i>f</i>)	2	0.02	61.99	<.0001	.00
<i>d</i> × <i>f</i>	2	0.02	52.36	<.0001	.00
<i>h</i> × <i>f</i>	4	0.13	191.81	<.0001	.01
<i>h</i> × <i>d</i>	2	0.06	166.02	<.0001	.01
<i>N</i> × <i>f</i>	6	0.00	1.98	0.065	.00
<i>N</i> × <i>d</i>	3	0.01	13.92	<.0001	.00
<i>N</i> × <i>h</i>	6	0.21	202.82	<.0001	.02
<i>h</i> × <i>d</i> × <i>f</i>	4	0.13	154.89	<.0001	.01
<i>N</i> × <i>d</i> × <i>f</i>	6	0.01	6.02	<.0001	.00
<i>N</i> × <i>h</i> × <i>d</i>	6	0.06	58.72	<.0001	.01
<i>N</i> × <i>h</i> × <i>f</i>	12	0.01	2.60	0.002	.00
<i>N</i> × <i>h</i> × <i>d</i> × <i>f</i>	12	0.01	4.07	<.0001	.00
Error	7128	1.25			

Note. *Prob.* = probability; $\hat{\omega}^2$ = estimated proportion of variance accounted for in the population by each effect.

were replicated. In both sets of reported results, the most important determinant of population factor recovery was communality level, regardless of the presence of model error. The other factors had little effect when communalities were high, and only when the communalities became low did N and overdetermination become important determinants of recovery of population factors. As discussed by MacCallum et al. (1999), these results show that with high communalities and strongly determined factors, sample size has relatively little impact on the solutions and good recovery of population factors can be achieved even with fairly small samples. However, even when the degree of overdetermination is strong, sample size has a much greater impact as communalities enter the wide or low range.

Of primary interest is the comparison of plots within Figures 1 and 2 in terms of model fit, or the comparison between the situations in which the common factor model has poor, fair, or very good fit in the population. Visual inspection of the mean plots reveals no major difference among fit levels. That is, comparison of panels 1, 3, and 5 within each figure, as well as comparison of panels 2, 4, and 6 within each figure, shows very little effect of the level of model error on these indexes of factor recovery. This observation is of course consistent with the ANOVA results indicating very small effect sizes for model fit as well as small interactions of model fit with other factors. Thus, results show no appreciable influence of model error on correspondence between sample and population factors, with this finding being consistent across levels of other design features considered here (sample size, communality level, and overdetermination). Finally, and importantly, levels of K in Figure 1 are very similar to corresponding levels in Figure 1 of MacCallum et al. (1999), where data contained no model error, providing further support for our hypothesis that correspondence between sample and population factors is not influenced by the presence or degree of model error.

No detailed results will be presented for the analysis of the bias measure, δ , since analyses showed essentially no bias in estimation of factor loadings under any of the conditions included in our study. The grand mean of δ was $-.006$, and the cell mean deviating *furthest* from zero was only $-.023$.

In summary, results of the Monte Carlo study support our hypotheses about the impact of model error on results of factor analysis of sample data. Effects of sample size, communality level, and overdetermination on recovery of population factors are essentially identical regardless of the presence of model error. These results indicate that conclusions regarding overdetermination, sample size, and level of communality drawn by MacCallum et al. (1999) should generalize to data for which the common factor model does not hold exactly. Results also showed essentially no bias

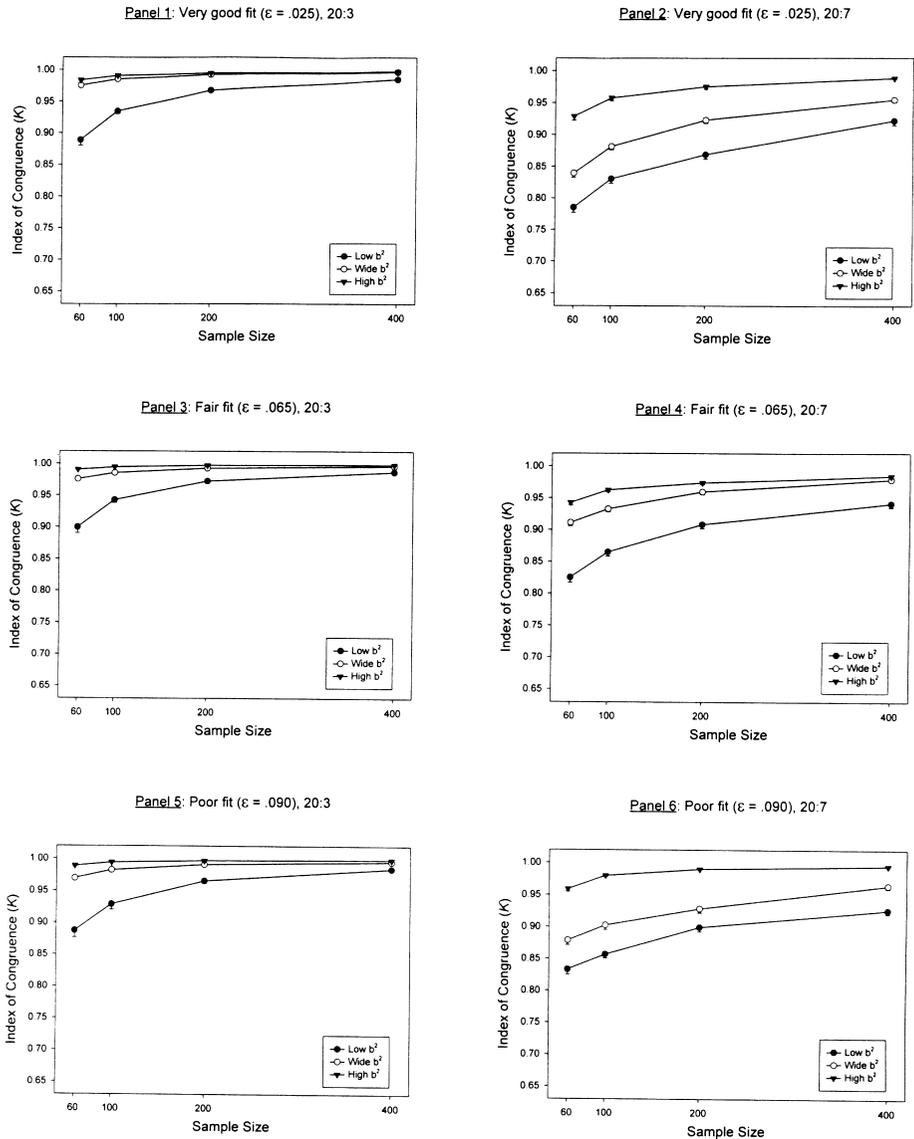


Figure 1

Cell Means for Index of Congruence (K) in Monte Carlo Study

Each panel represents a different combination of $p:r$ ratio and model fit. The vertical axis shows mean K between sample and population factors. Error bars show intervals within which 95% of the values of K lie for each cell (not visible when the width of the error bar is smaller than the symbol representing the mean). p = number of variables; r = number of factors.

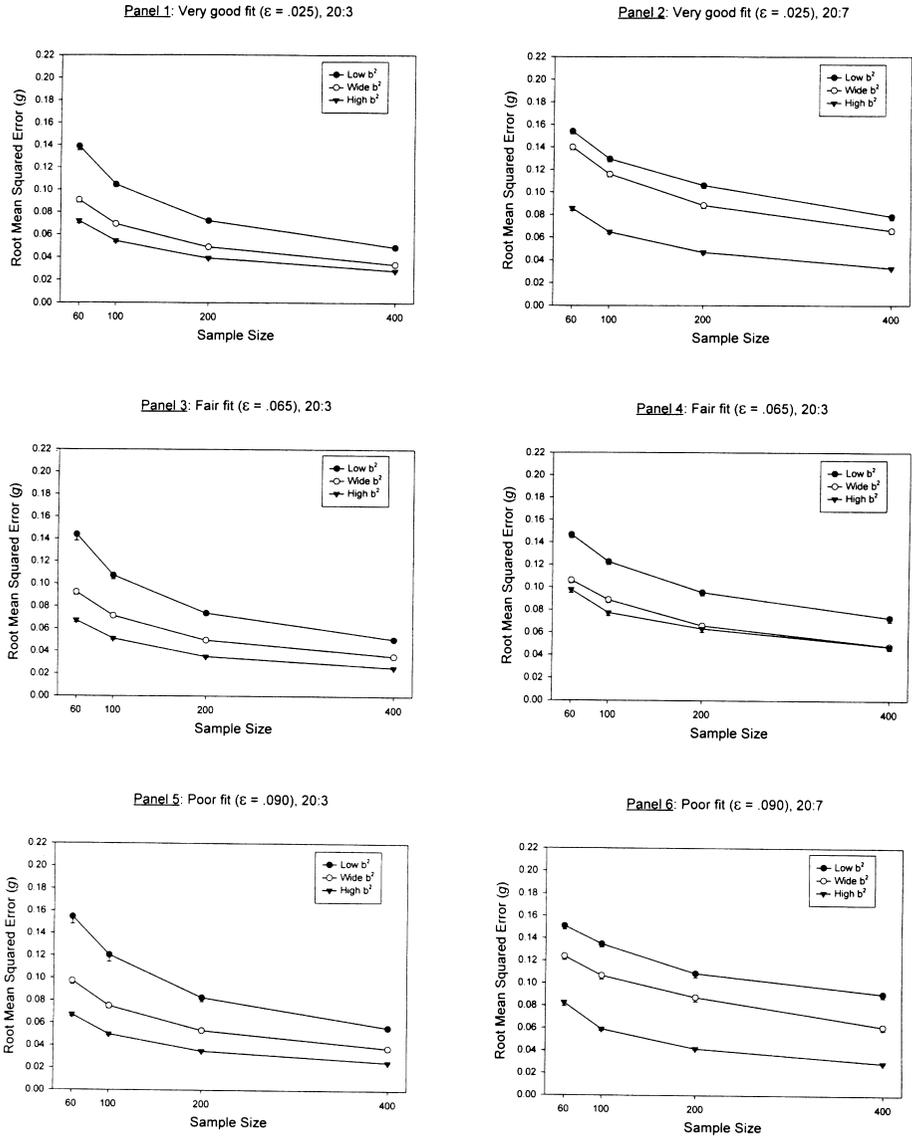


Figure 2
Cell Means for Index of Root Mean Squared Deviation (g) in Monte Carlo Study

in estimation of factor loadings. However, because this Monte Carlo study involved the generation of artificial data, we sought to demonstrate the validity of these conclusions in empirical data in which the nature of model error is not controlled.

Empirical Sampling Study

We conducted an empirical sampling study to determine whether the trends hypothesized from the theoretical framework presented earlier and verified in the Monte Carlo study would generalize to analyses of empirical data. We employed a data set obtained from the Department of Developmental Services (DDS) of the State of California containing measures on a 66-item inventory of adaptive behaviors for a large sample of mentally retarded or developmentally disabled persons.

Method

Samples. The total number of persons who were assessed in 1991 was 79,595. However, data from only those who were at least 15 years old in 1991 ($N = 55,085$) were used for this study because data from those who were less than age 15 might be less reliable. This group was considered to be a population from which samples were drawn as described below.

Instrument. The inventory of adaptive behaviors mandated by the California DDS is called the Client Development Evaluation Report (CDER). Widaman, Gibbs, and Geary (1987) found a highly stable six-factor structure for the CDER that was clearly replicated across 14 samples of persons with mental retardation. Widaman et al. labeled the six factors Motor Development (12 items), Independent Living Skills (10 items), Cognitive Competence (14 items), Social Competence (6 items), Social Maladaptation (10 items), and Personal Maladaptation (7 items).

Design. To investigate the issues of concern in the present study, we employed these empirical data and our knowledge of their factor structure to define various sets of measured variables that represented different levels of overdetermination and communalities. Two different levels of overdetermination were represented, as defined by the ratio $p:r$. These ratios were 12:3 and 18:3. More specific information on the selection of the variables and factors will be given later in this section.

In order to establish a method for manipulating level of communality, we performed preliminary factor analyses on selected sets of single items and other analyses on selected two-item parcels. Two-item parcels were formed by summing scores on two related items. Cattell and Burdsal (1975) showed

that the clarity and strength of the factorial description of parcels of items is greater than that of the constituent items if all items combined into a given parcel have the same factorial composition. As a result, the communality of a parcel should be higher than that of the individual items that comprise the parcel. Thus, two levels of communality were employed in the study: low (analyses based on items), and high (analyses based on two-item parcels).

Of the six CDER factors, three with a sufficient number of items (10 or more) were used for the present study; the three factors were Motor Development, Cognitive Competence, and Social Maladaptation. Items and parcels representing these three factors were selected so as to define data sets representing $p:r$ ratios of 12:3 and 18:3. For the low communality conditions, four items per factor were selected for the 12:3 condition of overdetermination, and six items per factor were selected for the 18:3 condition of overdetermination. For the high communality conditions at each $p:r$ level, items which had not been selected were paired with previously selected items, with resulting sums representing two-item parcels serving as indicators of the factors. Several items were used twice in forming two-item parcels to achieve high communalities.

A series of preliminary factor analyses conducted on the full population verified that the use of items and parcels did affect communality. Across the two solutions for which single items served as indicators, the mean communality of indicators was .58; for the two solutions having two-item parcels as indicators, the mean communality of indicators was .79. This manipulation of level of communality in the present empirical study must be considered small. In comparison to the Monte Carlo study presented earlier, note that the communalities in the "low" condition in the empirical sampling study (mean = .58) are not nearly as low as those in the low condition of the Monte Carlo study (range .2 to .4). As a result, effects of level of communality and its interactions in the empirical sampling study should be considered underestimates of potential effects on properties of factor analysis solutions.

For each combination of communality level (high, low) and level of overdetermination (18:3, 12:3), 100 random samples were drawn for each level of N . We used the same levels of N as in the simulation study: 60, 100, 200, and 400. Data from each of these samples were then factor analyzed using methods identical to those used in the preceding simulation study. As in the Monte Carlo study, samples resulting in Heywood cases were not screened out. (Frequency of Heywood cases was much lower in the empirical study due to the absence of a truly low-communality condition.) The congruence measures, K and g , and the bias measure, δ , as defined in the simulation study, were computed for each sample solution.

Evaluating Design Effects. In order to evaluate effects of sample size, level of communality, and level of overdetermination on quality of the factor solutions, the measures K , g , and δ were used as dependent variables in three-way ANOVAs. The ANOVA design was a 4 (levels of sample size) \times 2 (levels of communality) \times 2 (levels of overdetermination) design. It was hypothesized that results would follow the same pattern as in the Monte Carlo study, with one likely major difference: The effect of communality level would be less dramatic in the empirical study because the manipulation of this feature was smaller in the empirical study than in the simulation study. As in the simulation study, interpretation of ANOVA results will focus on effects showing substantial values of $\hat{\omega}^2$, regardless of statistical significance.

Results and Discussion

Let us first consider the fit of the common factor model in the empirical population of $N = 55,085$. Using maximum likelihood factor analysis, a three-factor model was fit to each of four population correlation matrices representing the four conditions of this design: communality level (high vs. low), and overdetermination (18:3 vs. 12:3). The resulting RMSEA values for each of these analyses were as follows: for high communality and $p:r = 18:3$, RMSEA = .099; for high communality and $p:r = 12:3$, RMSEA = .067; for low communality and $p:r = 18:3$, RMSEA = .090; for low communality and $p:r = 12:3$, RMSEA = .064. These results show higher levels of model error in those data sets where there were more indicators for the same three factors, regardless of communality level. This observation implies that adding more indicators may not always be beneficial because it may introduce additional model error, for instance in the form of more or stronger minor factors.

The ANOVA summary for the dependent variable K is presented in Table 3. The most prominent effects were the main effects of sample size ($\hat{\omega}^2 = .31$), communality level ($\hat{\omega}^2 = .52$), and overdetermination ($\hat{\omega}^2 = .01$). Effects involving communality level were much lower than in the Monte Carlo study because of the weaker manipulation of communality level in the empirical study. ANOVA summary results for g , which agree with the results for K , are presented in Table 4. To aid in interpretation of effects, cell means for K are presented in Figure 3 and those for g are presented in Figure 4. Effects of communality level and sample size followed the same pattern as observed in the Monte Carlo study. On average, good recovery of population factors was achieved under all conditions. Mean values of K were above .94 in all cells, and were above .99 for the high communality condition even when N was small. Recall that in the Monte Carlo study poor recovery

Table 3
ANOVA Results for the Measure of Congruence (*K*) in Empirical Sampling Study

Source	<i>df</i>	<i>SS</i>	<i>F</i>	<i>Prob.</i>	$\hat{\omega}^2$
Sample size (<i>N</i>)	3	155.14	1042.56	<.0001	.31
Communality (<i>h</i>)	1	262.33	5288.75	<.0001	.52
Overdetermination (<i>d</i>)	1	4.48	90.38	<.0001	.01
<i>h</i> × <i>d</i>	1	0.00	0.03	0.865	.00
<i>N</i> × <i>d</i>	3	0.11	0.75	0.521	.00
<i>N</i> × <i>h</i>	3	0.26	1.78	0.150	.00
<i>N</i> × <i>h</i> × <i>d</i>	3	0.09	0.61	0.606	.00
Error	1584	78.57			

Note. *Prob.* = probability; $\hat{\omega}^2$ = estimated proportion of variance accounted for in the population by each effect.

Table 4
ANOVA Results for the Root Mean Squared Error (*g*) in Empirical Sampling Study

Source	<i>df</i>	<i>SS</i>	<i>F</i>	<i>Prob.</i>	$\hat{\omega}^2$
Sample size (<i>N</i>)	3	0.70	652.21	<.0001	.33
Communality (<i>h</i>)	1	0.76	2116.97	<.0001	.36
Overdetermination (<i>d</i>)	1	0.01	32.35	<.0001	.01
<i>h</i> × <i>d</i>	1	0.00	2.42	0.120	.00
<i>N</i> × <i>d</i>	3	0.00	0.05	0.984	.00
<i>N</i> × <i>h</i>	3	0.09	80.52	0.000	.04
<i>N</i> × <i>h</i> × <i>d</i>	3	0.00	0.22	0.881	.00
Error	1584	0.57			

Note. *Prob.* = probability; $\hat{\omega}^2$ = estimated proportion of variance accounted for in the population by each effect.

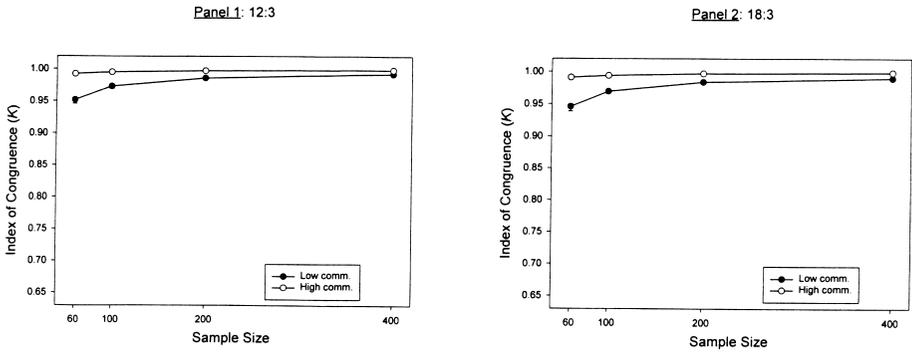


Figure 3
Cell Means for Index of Congruence (K) in Empirical Sampling Study

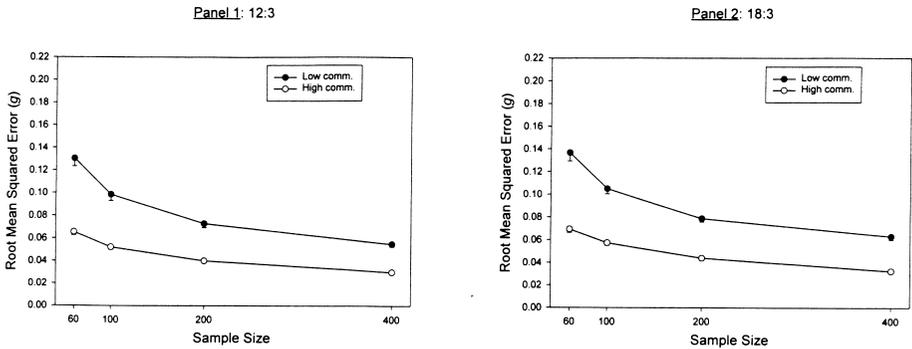


Figure 4
Cell Means for Root Mean Squared Deviation (g) in Monte Carlo Study

was observed when communalities were low. That apparent discrepancy in results is illusory, however, because the empirical study contained no comparable low-communality condition. The low-communality condition in the empirical study is more comparable to the wide-communality condition in the Monte Carlo study.

The effect of level of overdetermination, as well as its interactive effect with level of communality, was found to be much weaker in the empirical study than in the Monte Carlo study. The lack of a significant interaction between overdetermination and communality level can probably be

attributed to the weaker manipulation of communality in the empirical study. The difference between the main effects of overdetermination in the two studies may be attributable to the difference in the way in which this feature was manipulated. In the Monte Carlo study, levels of overdetermination were represented by $p:r$ ratios of 20:3 and 20:7. That is, the number of indicators was held constant while the number of factors was varied. In the empirical study, levels of overdetermination were represented by $p:r$ ratios of 18:3 and 12:3. The number and nature of the three factors were held constant, and the number of indicators was varied. Given that the three factors in the empirical study were well represented by the 12 indicators, the addition of two more indicators per factor did not result in a substantial change in recovery of those factors. By comparison, the Monte Carlo results showed a more dramatic effect in attempts to recover seven vs. three factors using a fixed number of indicators ($p = 20$). This finding of a difference in the effect of overdetermination depending on whether p or r is varied is consistent with speculations by MacCallum et al. (1999, p. 90) based on their theoretical framework.

From another perspective, the small influence of overdetermination on K and g lends further support to our primary hypothesis that recovery of factors is unaffected by model error. Recall that factor analyses of the four "population" correlation matrices for the empirical study showed poorer fit for the matrices involving 18 variables than for those involving 12 variables. Importantly, results of analysis of K and g show that correspondence between sample and population factors was not appreciably different in those conditions, despite the difference in model fit.

Results for analysis of the bias measure, δ , will not be presented in detail. As in the Monte Carlo study, no evidence of bias was found, with mean values of δ being near zero in every cell.

In summary, results of the empirical sampling study verify that the phenomena observed in the Monte Carlo study generalize to empirical data. Most importantly, as long as communalities are high and factors are adequately overdetermined, population factors will, on the average, be recovered accurately in analysis of sample data even when N is relatively small. Furthermore, correspondence between sample and population factors does not appear to be influenced by the degree of model error.

General Discussion

In this article we have examined the issue of sample size in factor analysis in the presence of model error. This study extends recent work by MacCallum et al. (1999) and Velicer and Fava (1998), as well as earlier

studies of this issue, which made use of the context where no model error was present. While theoretically interesting, such a perspective is empirically unrealistic. Our current approach made use of a theoretical framework presented by MacCallum and Tucker (1991) to represent model error explicitly in the population and sample and to differentiate the roles of model error and sampling error. This approach provided a basis for predicting influences of model error, sample size, and other aspects of design and data on recovery of population factors in sample data. The primary hypothesis was that, on average, recovery of population factors would not be adversely affected by the presence of model error. This prediction was verified in our simulation and empirical sampling studies. Results showed that effects of sample size, communality level, and level of overdetermination of factors on recovery of population factors were essentially unaltered by the presence of model error.

A limitation of the current study is that all factor analyses were conducted using the known correct number of factors. As has been established in previous studies (e.g., Fava & Velicer, 1992, 1996), retention of an incorrect number of factors, especially retention of too few factors, can cause major distortion of loading patterns. Although underfactoring can be considered as a source of major model error, our results regarding lack of effects of model error on factor recovery do not generalize to such a situation. Our findings pertain primarily to influences of an aggregated variety of relatively minor sources of model error, rather than the type of major misspecification represented by failure to retain a major common factor. Given the serious consequences of underfactoring, it is clear that users of factor analysis should be very rigorous in making the number-of-factors decision and should err in the direction of overfactoring when the evidence is ambiguous.

Certainly other limitations apply to our Monte Carlo study, as is the case with any such study. The range of conditions considered and the nature of the simulated data cannot fully represent the variety of conditions and data structures encountered in empirical research. Given these limitations, it is important to keep in mind that the hypotheses based on our theoretical framework and confirmed in our Monte Carlo study were also strongly supported in our sampling study using empirical data.

It is important to consider how our results should be viewed relative to the use of exploratory versus confirmatory factor analysis. On one hand, our Monte Carlo and empirical sampling studies involved exploratory factor analysis. However, our theoretical framework involved the general common factor model, and the resulting hypotheses should be equally valid in both exploratory and confirmatory settings. Although we have shown the

hypothesized effects to be present in exploratory analyses, testing these same effects in confirmatory factor analysis would require careful consideration of some issues that distinguish exploratory from confirmatory approaches. Especially relevant in this context would be the nature of model error. The increased potential for misspecification in confirmatory analyses introduces additional and different kinds of model error beyond those evaluated in our sampling studies. The validity of our hypotheses in the presence of such model error may be a topic for further study. We are confident, though, that the effects described in the current article should generalize from exploratory to confirmatory analyses as long as the confirmatory model is not badly misspecified.

Within the context of exploratory factor analysis, our results have more relevance than do previous results for empirical studies because of the explicit incorporation of model error, which will always be present to some degree in practice. Our results also reinforce the evidence regarding the critical role of communality level, while demonstrating the limited value of traditional rules of thumb regarding sample size in factor analysis. For example, one common rule of thumb regarding sample size is that, to ensure stability of a factor solution, a researcher should have a ratio of subjects to variables that is 4:1 or larger. However, our results show that if communalities are high, recovery of population factors in sample data is normally very good, almost regardless of sample size, level of overdetermination, or the presence of model error. Thus, samples somewhat smaller than traditionally recommended are likely sufficient when communalities are high. When communalities are lower, much larger samples are needed, with this phenomenon being amplified by poorly overdetermined factors. Indeed, sample sizes may have to be much larger than typically recommended (e.g., a 20:1 subjects-to-variables ratio) when communalities are low and factors are not highly overdetermined. Our theoretical framework along with results of sampling studies, in combination with those of MacCallum et al. (1999), provide a more informed view of the issue of sample size in factor analysis. It is now clear that it is not possible to make blanket recommendations regarding this issue without considering other important aspects of design.

References

- Arrindell, W. A. & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, 9, 165-178.
- Barrett, P. T. & Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality Study in Group Behavior*, 1, 23-33.

- Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika*, *33*, 267-334.
- Browne, M. W. & 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, CA: Sage.
- Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum.
- Cattell, R. B. & Burdsal, C. A., Jr. (1975). The radial parcel double factorial design: A solution to the item vs. parcel controversy. *Multivariate Behavioral Research*, *10*, 165-179.
- Comrey, A. L. & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.
- Cudeck, R. & Henly, S. J. (1991). Model selection in covariance structure analysis and the "problem" of sample size: A clarification. *Psychological Bulletin*, *109*, 512-519.
- Fava, J. L. & Velicer, W. F. (1992). 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 analysis. *Educational and Psychological Measurement*, *56*, 907-929.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- 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.
- Maxwell, S. E. & Delaney, H. D. (1990). *Designing experiments and analyzing data*. Belmont, CA: Wadsworth.
- Pennell, R. (1968). The influence of communality and N on the sampling distributions of factor loadings. *Psychometrika*, *33*, 423-439.
- Steiger, J. H. (1989). *Causal modeling: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT.
- Steiger, J. H. & Lind, J. M. (1980, June). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Tateneni, K., Mels, G., Cudeck, R., & Browne, M. W. (1998). *CEFA: Comprehensive exploratory factor analysis* [Computer software]. Columbus, OH: Authors.
- 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. & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*, 231-251.
- 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., Gibbs, K. W., & Geary, D. C. (1987). The structure of adaptive behavior: I. Replication across fourteen samples of nonprofoundly mentally retarded people. *American Journal of Mental Deficiency*, *91*, 348-360.
- Wijisman, R. A. (1959). Applications of a certain representation of the Wishart matrix. *Annals of Mathematical Statistics*, *30*, 597-601.

Accepted December, 2000.