

## Sample Size in Factor Analysis

Robert C. MacCallum  
Ohio State University

Keith F. Widaman  
University of California, Riverside

Shaobo Zhang and Sehee Hong  
Ohio State University

The factor analysis literature includes a range of recommendations regarding the minimum sample size necessary to obtain factor solutions that are adequately stable and that correspond closely to population factors. A fundamental misconception about this issue is that the minimum sample size, or the minimum ratio of sample size to the number of variables, is invariant across studies. In fact, necessary sample size is dependent on several aspects of any given study, including the level of communality of the variables and the level of overdetermination of the factors. The authors present a theoretical and mathematical framework that provides a basis for understanding and predicting these effects. The hypothesized effects are verified by a sampling study using artificial data. Results demonstrate the lack of validity of common rules of thumb and provide a basis for establishing guidelines for sample size in factor analysis.

In the factor analysis literature, much attention has been given to the issue of sample size. It is widely understood that the use of larger samples in applications of factor analysis tends to provide results such that sample factor loadings are more precise estimates of population loadings and are also more stable, or less variable, across repeated sampling. Despite general agreement on this matter, there is considerable divergence of opinion and evidence about the question of how large a sample is necessary to adequately achieve these objectives. Recommendations and findings about this issue are diverse and often contradictory. The objectives of this article are to provide a

clear understanding and demonstration of how sample size influences factor analysis solutions and to offer useful and well-supported recommendations on desirable sample sizes for empirical research. In the process, we also hope to account for some inconsistent findings and recommendations in the literature.

A wide range of recommendations regarding sample size in factor analysis has been proposed. These guidelines typically are stated in terms of either the minimum necessary sample size,  $N$ , or the minimum ratio of  $N$  to the number of variables being analyzed,  $p$ . Many of these guidelines were reviewed and discussed by Arrindell and van der Ende (1985) and more recently by Velicer and Fava (1998). Let us consider a sampling of recommendations regarding absolute sample size. Gorsuch (1983) recommended that  $N$  should be at least 100, and Kline (1979) supported this recommendation. Guilford (1954) argued that  $N$  should be at least 200, and Cattell (1978) claimed the minimum desirable  $N$  to be 250. Comrey and Lee (1992) offered a rough rating scale for adequate sample sizes in factor analysis: 100 = *poor*, 200 = *fair*, 300 = *good*, 500 = *very good*, 1,000 or more = *excellent*. They urged researchers to obtain samples of 500 or more observations whenever possible in factor analytic studies.

Considering recommendations for the  $N:p$  ratio, Cattell (1978) believed this ratio should be in the

---

Robert C. MacCallum, Shaobo Zhang, and Sehee Hong, Department of Psychology, Ohio State University; Keith F. Widaman, Department of Psychology, University of California, Riverside. Sehee Hong is now at the Department of Education, University of California, Santa Barbara.

We acknowledge the contributions and insights of Ledyard Tucker during the development of the theory and procedures presented in this article. We thank Michael Browne for comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Robert C. MacCallum, Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, Ohio 43210-1222. Electronic mail may be sent to maccallum.1@osu.edu.

range of 3 to 6. Gorsuch (1983) argued for a minimum ratio of 5. Everitt (1975) recommended that the  $N:p$  ratio should be at least 10. Clearly the wide range in these recommendations causes them to be of rather limited value to empirical researchers. The inconsistency in the recommendations probably can be attributed, at least in part, to the relatively small amount of explicit evidence or support that is provided for any of them. Rather, most of them seem to represent general guidelines developed from substantial experience on the part of their supporters. Some authors (e.g., Comrey & Lee, 1992) placed the sample size question into the context of the need to make standard errors of correlation coefficients adequately small so that ensuing factor analyses of those correlations would yield stable solutions. It is interesting that a number of important references on factor analysis make no explicit recommendation at all about sample size. These include books by Harman (1976), Lawley and Maxwell (1971), McDonald (1985), and Mulaik (1972).

There are some research findings that are relevant to the sample size question. There is considerable literature, for instance, on the topic of standard errors in factor loadings. As sample size increases, the variability in factor loadings across repeated samples will decrease (i.e., the loadings will have smaller standard errors). Formulas for estimating standard errors of factor loadings have been developed for various types of unrotated loadings (Girshick, 1939; Jennrich, 1974; Lawley & Maxwell, 1971) and rotated loadings (Archer & Jennrich, 1973; Jennrich, 1973, 1974). Cudeck and O'Dell (1994) reviewed the literature on this subject and provided further developments for obtaining standard errors of rotated factor loadings. Application of these methods would demonstrate that standard errors decrease as  $N$  increases. In addition, Archer and Jennrich (1976) showed this effect in a Monte Carlo study. Although this effect is well-defined theoretically and has been demonstrated with simulations, there is no guidance available to indicate how large  $N$  must be to obtain adequately small standard errors of loadings. A detailed study of this question would be difficult because, as emphasized by Cudeck and O'Dell, these standard errors depend in a complex way on many things other than sample size, including method of rotation, number of factors, and the degree of correlation among the factors. Furthermore, a general method for obtaining standard errors for rotated loadings has only recently been developed (Browne & Cudeck, 1997).

Some Monte Carlo studies have demonstrated ef-

fects of sample size on factor solutions. For instance, Browne (1968) investigated the quality of solutions produced by different factor analytic methods. Browne found that solutions obtained from larger samples showed greater stability (smaller standard deviations of loadings across repeated samples) and more accurate recovery of the population loadings. Browne also found that better recovery of population solutions was obtained when the ratio of the number of variables,  $p$ , to the number of factors,  $r$ , increased. This issue is discussed further in the next section and is of major interest in this article. In another early study, Pennell (1968) examined effects of sample size on stability of loadings and found that effect to diminish as communalities of the variables increased. (The communality of a variable is the portion of the variance of that variable that is accounted for by the common factors.) In a study comparing factoring methods, Velicer, Peacock, and Jackson (1982) found similar effects. They observed that recovery of true structure improved with larger  $N$  and with higher saturation of the factors (i.e., when the nonzero loadings on the factors were higher). These same trends were replicated in a recent study by Velicer and Fava (1998), who also found the influence of sample size to be reduced when factor loadings, and thus communalities, were higher. Although these studies demonstrated clear effects of sample size, as well as some evidence for an interactive effect of sample size and level of communality, they did not shed much light on the issue of establishing a minimum desirable level of  $N$ , nor did they offer an integrated theoretical framework that accounted for these various effects.

Additional relevant information is reported in two empirical studies of the sample size question. Barrett and Kline (1981) investigated this issue by using two large empirical data sets. One set consisted of measures for 491 participants on the 16 scales of Cattell's Sixteen Personality Factor Questionnaire (16 PF; Cattell, Eber, & Tatsuoka, 1970). The second set consisted of measures for 1,198 participants on the 90 items of the Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975). From each data set, Barrett and Kline drew subsamples of various sizes and carried out factor analyses. They compared the subsample solutions with solutions obtained from the corresponding full samples and found that good recovery of the full-sample solutions was obtained with relatively small subsamples. For the 16 PF data, good recovery was obtained from a subsample of  $N = 48$ . For the EPQ data, good recovery was obtained from a

subsample of  $N = 112$ . These values represent  $N:p$  ratios of 3.0 and 1.2, respectively. These findings represent much smaller levels of  $N$  and  $N:p$  than those recommended in the literature reviewed above. Barrett and Kline suggested that the level of recovery they achieved with rather small samples might have been a function of data sets characterized by strong, clear factors, a point we consider later in this article.

A similar study was conducted by Arrindell and van der Ende (1985). These researchers used large samples of observations on two different fear questionnaires and followed procedures similar to Barrett and Kline's (1981) to draw subsamples and compare their solutions with the full-sample solutions. For a 76-item questionnaire, they found a subsample of  $N = 100$  sufficient to achieve an adequate match to the full-sample solution. For a 20-item questionnaire, they found a subsample of  $N = 78$  sufficient. These  $N$ s correspond to  $N:p$  ratios of 1.3 and 3.9, respectively. Again, these results suggest that samples smaller than generally recommended might be adequate in some applied factor analysis studies.

We suggest that previous recommendations regarding the issue of sample size in factor analysis have been based on a misconception. That misconception is that the minimum level of  $N$  (or the minimum  $N:p$  ratio) to achieve adequate stability and recovery of population factors is invariant across studies. We show that such a view is incorrect and that the necessary  $N$  is in fact highly dependent on several specific aspects of a given study. Under some conditions, relatively small samples may be entirely adequate, whereas under other conditions, very large samples may be inadequate. We identify aspects of a study that influence the necessary sample size. This is first done in a theoretical framework and is then verified and investigated further with simulated data. On the basis of these findings we provide guidelines that can be used in practice to estimate the necessary sample size in an empirical study.

### Influences of Sample Size on Factor Analysis Solutions

MacCallum and Tucker (1991) presented a theoretical analysis of sources of error in factor analysis, focusing in part on how random sampling influences parameter estimates and model fit. Their framework provides a basis for understanding and predicting the influence of sample size in factor analysis. MacCallum and Tucker distinguished between "model error,"

arising from lack of fit of the model in the population, and "sampling error," arising from lack of exact correspondence between a sample and a population. This distinction is essentially equivalent to that drawn by Cudeck and Henly (1991) between "discrepancy of approximation" and "discrepancy of estimation." MacCallum and Tucker isolated various effects of model error and sampling error in the common factor model. The influence of sampling error is to introduce inaccuracy and variability in parameter estimates, whereas the influence of model error is to introduce lack of fit of the model in the population and the sample. In any given sample, these two phenomena act in concert to produce lack of model fit and error in parameter estimates. In the following developments, because we are focusing on the sample size question, we eliminate model error from consideration and make use of a theoretical framework wherein the common factor model is assumed to hold exactly in the population. The following framework is adapted from MacCallum and Tucker, given that condition. Readers familiar with factor analytic theory will note that parts of this framework are nonstandard representations of the model. However, this approach provides a basis for important insights about the effects of the role of sample size in factor analysis. The following mathematical developments make use of simple matrix algebra. Resulting hypotheses are summarized in the *Summary of Hypotheses* section.

### Mathematical Framework

Let  $\mathbf{y}$  be a random row vector of scores on  $p$  measured variables. (In all of the following developments, it is assumed that all variables, both observed and latent, have means of zero.) The common factor model can be expressed as

$$\mathbf{y} = \mathbf{x}\Omega', \quad (1)$$

where  $\mathbf{x}$  is a row vector of scores on  $r$  common and  $p$  unique factors, and  $\Omega$  is a matrix of population factor loadings for common and unique factors. The vector  $\mathbf{x}$  can be partitioned as

$$\mathbf{x} = [\mathbf{x}_c, \mathbf{x}_u], \quad (2)$$

where  $\mathbf{x}_c$  contains scores on  $r$  common factors, and  $\mathbf{x}_u$  contains scores on  $p$  unique factors. Similarly, the factor loading matrix  $\Omega$  can be partitioned as

$$\Omega = [\Lambda, \Theta]. \quad (3)$$

Matrix  $\Lambda$  is of order  $p \times r$  and has entries representing population common factor loadings. Matrix  $\Theta$  is a

diagonal matrix of order  $p \times p$  and has diagonal entries representing unique factor loadings.

Equation 1 expresses measured variables as linear combinations of common and unique factors. Following a well-known expression for variances and covariances of linear combinations of random variables (e.g., Morrison, 1990, p. 84), we can write an expression for the population covariance matrix of the measured variables,  $\Sigma_{yy}$ :

$$\Sigma_{yy} = \Omega \Sigma_{xx} \Omega', \quad (4)$$

where  $\Sigma_{xx}$  is the population covariance matrix for the common and unique factor scores in vector  $\mathbf{x}$ . Matrix  $\Sigma_{xx}$  is a square matrix of order  $(r + p)$  and can be viewed as being partitioned into sections as follows:

$$\Sigma_{xx} = \begin{bmatrix} \Sigma_{cc} & \Sigma_{cu} \\ \Sigma_{uc} & \Sigma_{uu} \end{bmatrix}. \quad (5)$$

Each submatrix of  $\Sigma_{xx}$  corresponds to a population covariance matrix for certain types of factors. Matrix  $\Sigma_{cc}$  is the  $r \times r$  population covariance matrix for the common factors. Matrix  $\Sigma_{uu}$  is the  $p \times p$  population covariance matrix for the unique factors. Matrices  $\Sigma_{uc}$  and  $\Sigma_{cu}$  contain population covariances of common factors with unique factors, where  $\Sigma_{uc}$  is the transpose of  $\Sigma_{cu}$ . Without loss of generality, we can define all of the factors as being standardized in the population. This allows us to view  $\Sigma_{xx}$  and each of its submatrices as containing correlations. Given this specification, let us define a matrix  $\Phi$  whose entries contain population correlations among the common factors. That is,

$$\Sigma_{cc} = \Phi. \quad (6)$$

Furthermore, because unique factors must, by definition, be uncorrelated with each other and with common factors in the population, a number of simplifications of  $\Sigma_{xx}$  can be recognized. In particular, matrix  $\Sigma_{uu}$  must be an identity matrix ( $I$ ), and matrices  $\Sigma_{uc}$  and  $\Sigma_{cu}$  must contain all zeros:

$$\Sigma_{uu} = I. \quad (7)$$

$$\Sigma_{uc} = \Sigma'_{cu} = 0. \quad (8)$$

Given these definitions, we can rewrite Equation 5 as follows:

$$\Sigma_{xx} = \begin{bmatrix} \Phi & 0 \\ 0 & I \end{bmatrix}. \quad (9)$$

Substituting from Equation 9 into Equation 4 and expanding, we obtain the following:

$$\Sigma_{yy} = \Lambda \Phi \Lambda' + \Theta^2. \quad (10)$$

This is a fundamental expression of the common factor model in the population. It defines the population covariances for the measured variables as a function of the parameters in  $\Lambda$ ,  $\Phi$ , and  $\Theta^2$ . Note that the entries in  $\Theta^2$  are squared unique factor loadings, also called unique variances. These values represent the variance in each variable that is not accounted for by the common factors.

Equation 10 represents a standard expression of the common factor model in the population. If the model in Equation 1 is true, then the model for the population covariance matrix in Equation 10 will hold exactly. Let us now consider the structure of a *sample* covariance matrix as implied by the common factor model. A sample covariance matrix differs from the population covariance matrix because of sampling error; as a result, the structure in Equation 10 will not hold exactly in a sample. To achieve an understanding of the nature of sampling error in this context, it is desirable to formally represent the cause of such lack of fit. To do so, we must consider what aspects of the model are the same in a sample as they are in a population, as well as what aspects are not the same. Let us first consider factor loadings. According to common factor theory, the common and unique factor loadings are weights that are applied to the common and unique factors in a linear equation to approximate the observed variables, as in Equation 1. These weights are the same for each individual. Thus, the *true* factor loadings in every sample will be the same as they are in the population. (This is not to say that the sample *estimates* of those loadings will be exact or the same in every sample. They will not, for reasons to be explained shortly. Rather, the true loadings are the same in every sample.) The matrix  $\Omega$  then represents the true factor loading matrix for the sample as well as the population.

Next consider the relationships among the factors. The covariances among the factors will generally not be the same in a sample as they are in the population. This point can be understood as follows. Every individual in the population has (unknowable) scores on the common and unique factors. Matrix  $\Sigma_{xx}$  is the covariance matrix of those scores. If we were to draw a sample from the population, and if we knew the common and unique factor scores for the individuals in that sample, we could compute a *sample* covariance matrix  $C_{xx}$  for those scores. The elements in  $C_{xx}$  would differ from those in  $\Sigma_{xx}$  simply because of random sampling fluctuation, in the same manner that sample variances and covariances of measured vari-

ables differ from corresponding population values. Matrix  $C_{xx}$  can be viewed as being made up of submatrices as follows:

$$C_{xx} = \begin{bmatrix} C_{cc} & C_{cu} \\ C_{uc} & C_{uu} \end{bmatrix}, \quad (11)$$

where the submatrices contain specified sample covariances among common and unique factors.

Thus, we see that true factor loadings are the same in a sample as in the population, but that factor covariances are not the same. This point provides a basis for understanding the factorial structure of a sample covariance matrix. Earlier we explained that the population form of the common factor model in Equation 4 was based on a general expression for variances and covariances of random variables (e.g., Morrison, 1990, p. 84). That same general rule applies in a sample as well. If we were to draw a sample of individuals for whom the model in Equation 1 holds and were to compute the sample covariance matrix  $C_{yy}$  for the measured variables, that matrix would have the following structure:

$$C_{yy} = \Omega C_{xx} \Omega'. \quad (12)$$

Substituting from Equations 3 and 11 into Equation 12 yields the following:

$$C_{yy} = \Lambda C_{cc} \Lambda' + \Lambda C_{cu} \Theta' + \Theta C_{uc} \Lambda' + \Theta C_{uu} \Theta'. \quad (13)$$

This equation represents a full expression of the structure of a sample covariance matrix under the assumption that the common factor model in Equation 1 holds exactly. That is, if Equation 1 holds for all individuals in a population, and consequently for all individuals in any sample from the population, then Equation 13 holds in the sample. Alternatively, if the standard factor analytic model in Equation 10 is exactly true in the population, then the structure in Equation 13 is exactly true in any sample from that population. Interestingly, Equation 13 shows that the *sample* covariances of the measured variables may be expressed as a function of *population* common and unique loadings (in  $\Lambda$  and  $\Theta$ ) and *sample* covariances for common and unique factors (in  $C_{cc}$ ,  $C_{cu}$ ,  $C_{uc}$ , and  $C_{uu}$ ).<sup>1</sup> As we discuss in the next section, this view provides insights about the impact of sample size in factor analysis.

In an ideal world, we could use Equation 13 to obtain exact values of population factor loadings. If we knew the factor scores for all individuals in a sample, we would then know the sample covariance

matrices for the common and unique factors. If we were to then solve Equation 13 for  $\Lambda$  and  $\Theta$ , that solution would exactly correspond to the population common and unique factor loadings. Of course, this procedure cannot be followed in practice because we do not know the factor scores. Thus, the structure represented in Equation 13 cannot be used to conduct factor analysis in a sample.

Rather, factor analysis in a sample is conducted using the population model in Equation 10. That model is fit to a sample covariance matrix,  $C_{yy}$ , yielding estimates of  $\Lambda$ ,  $\Phi$ , and  $\Theta$ , along with information about degree of fit of the model. The current framework provides a basis for understanding one primary source of such lack of fit. That source involves the relationships among the common and unique factors. When the model in Equation 10 is fit to sample data, it is assumed that unique factors are uncorrelated with each other and with common factors. Although this is true in the population, it will generally not be true in a sample. This phenomenon manifests itself in the difference between Equations 10 and 13. Specifically, unique factors will exhibit nonzero covariances with each other and with common factors in a sample, simply because of sampling error. This perspective provides a basis for understanding how sampling error influences the estimation of factor loadings. Equation 13 is key. Consider first the ideal case where the variances and covariances in a sample exactly match the corresponding population values (i.e.,  $C_{cc} = \Phi$ ;  $C_{cu} = 0$ ;  $C_{uc} = 0$ ;  $C_{uu} = I$ ). In that case, Equation 13 reduces to

$$C_{yy} = \Lambda \Phi \Lambda' + \Theta^2. \quad (14)$$

That is, the population model in Equation 10 would hold exactly in such a sample. In such an ideal case, a factor analysis of  $C_{yy}$  would yield an exact solution for the population parameters of the factor model, regardless of the size of the sample. If unique factors were truly uncorrelated with each other and with common factors in a sample, and if sample correlations among common factors exactly matched correspond-

<sup>1</sup> This is a nonstandard representation of the factor structure of a sample covariance matrix but is not original in the current work. For example, Cliff and Hamburger (1967) provided an equation (their Equation 4, p. 432) that is essentially equivalent to our Equation 12, which is the basis for Equation 13, and pointed out the mixture of population and sample values in this structure.

ing population values, then factor analysis of sample data would yield factor loadings that were exactly equal to population values. In reality, however, such conditions would not hold in a sample. Most importantly,  $C_{uu}$  would not be diagonal, and  $C_{cu}$  and  $C_{uc}$  would not be zero. Thus, Equation 14 would not hold. We would then fit such a model to  $C_{yy}$ , and the resulting solution would not be exact but would yield estimates of model parameters. Thus, the lack of fit of the model to  $C_{yy}$  can be seen as being attributable to random sampling fluctuation in the variances and covariances of the factors, with the primary source of error being the presence of nonzero sample covariances of unique factors with each other and with common factors. The further these covariances deviate from zero, the greater the effect on the estimates of model parameters and on fit of the model.

### *Implications for Effects of Sample Size*

This framework provides a basis for understanding some effects of sample size in fitting the common factor model to sample data. The most obvious sampling effect arises from nonzero sample intercorrelations of unique factors with each other and with common factors. Note that as  $N$  increases, these intercorrelations will tend to approach their population values of zero. As this occurs, the structure in Equation 13 will correspond more closely to that in Equation 14, and the lack of fit of the model will be reduced. Thus, as  $N$  increases, the assumption that the intercorrelations of unique factors with each other and with common factors will be zero in the sample will hold more closely, thereby reducing the impact of this type of sampling error and, in turn, improving the accuracy of parameter estimates.<sup>2</sup>

A second effect of sampling involves the magnitude of the unique factor loadings in  $\Theta$ . Note that in Equation 13 these loadings in effect function as weights for the elements in  $C_{cu}$ ,  $C_{uc}$ , and  $C_{uu}$ . Suppose the communalities of all measured variables are high, meaning that unique factor loadings are low. In such a case, the entries in  $C_{cu}$ ,  $C_{uc}$ , and  $C_{uu}$  receive low weight and in turn contribute little to the structure of  $C_{yy}$ . Thus, even if unique factors exhibited nontrivial sample covariances with each other and with common factors, the impact of that source of sampling error would be greatly reduced by the role of the low unique factor loadings in  $\Theta$ . Under such conditions, the influence of this source of sampling error would be minor, regardless of sample size. On the other hand, if communalities were low, meaning unique

factor loadings in  $\Theta$  were high, then the elements of  $C_{cu}$ ,  $C_{uc}$ , and  $C_{uu}$  would receive more weight in Equation 13 and would make a larger contribution to the structure of  $C_{yy}$  and in turn to lack of fit of the model. In that case, the role of sample size becomes more important. In smaller samples, the elements of  $C_{cu}$  and  $C_{uc}$  and the off-diagonals of  $C_{uu}$  would tend to deviate further from zero. If those matrices receive higher weight from the diagonals of  $\Theta$ , then the impact of sample size increases.

These observations can be translated into a statement regarding the expected influence of unique factor weights on recovery of population factors in analysis of sample data. We expect both a main effect of unique factor weights and an interactive effect between those weights and sample size. When unique factor weights are small (high communalities), the impact of this source of sampling error will be small regardless of sample size, and recovery of population factors should be good. However, as unique factor weights become larger (low communalities), the impact of this source of sampling error is more strongly influenced by sample size, causing poorer recovery of population factors. The effect of communality size on quality of sample factor solutions (replicability, recovery of population factors) has been noted previously in the literature on Monte Carlo studies (e.g., Cliff & Hamburger, 1967; Cliff & Pennell, 1967; Pennell, 1968; Velicer et al., 1982; Velicer & Fava, 1998), but the theoretical basis for this phenomenon, as presented above, has not been widely recognized. Gorsuch (1983, p. 331) noted in referring to an equation similar to our Equation 13, that unique factor weights are applied to the sample correlations involving unique factors and will thus impact replicability of factor solutions. However, Gorsuch did not discuss the interaction between this effect and sample size.

Finally, we consider an additional aspect of factor analysis studies and its potential interaction with sample size. An important issue affecting the quality of factor analysis solutions is the degree of overdetermination of the common factors. By overdetermination we mean the degree to which each factor is clearly represented by a sufficient number of variables. An important component of overdetermination

<sup>2</sup> Gorsuch (1983, p. 331) mentioned that replicability of factor solutions will be enhanced when correlations involving unique factors are low, but he did not explicitly discuss the role of sample size in this phenomenon.

is the ratio of the number of variables to the number of factors,  $p:r$ , although overdetermination is not defined purely by this ratio. Highly overdetermined factors are those that exhibit high loadings on a substantial number of variables (at least three or four) as well as good simple structure. Weakly overdetermined factors tend to exhibit poor simple structure without a substantial number of high loadings. Although overdetermination is a complex concept, the  $p:r$  ratio is an important aspect of it that has received attention in the literature. For all factors to be highly overdetermined, it is desirable that the number of variables be at least several times the number of factors. Comrey and Lee (1992, pp. 206–209) discussed overdetermination and recommended at least five times as many variables as factors. This is not a necessary condition for a successful factor analysis and does not assure highly overdetermined factors, but rather it is an aspect of design that is desirable to achieve both clear simple structure and highly overdetermined factors. In a major simulation study, Tucker, Koopman, and Linn (1969) demonstrated that several aspects of factor analysis solutions are much improved when the ratio of the number of variables to the number of factors increases. This effect has also been observed by Browne (1968) and by Velicer and Fava (1987, 1998).

With respect to sample size, it has been speculated (Arrindell & van der Ende, 1985; Barrett & Kline, 1981) that the impact of sampling error on factor analytic solutions may be reduced when factors are highly overdetermined. According to these investigators, there may exist an interactive effect between sample size and degree of overdetermination of the factors, such that when factors are highly overdetermined, sample size may have less impact on the quality of results. Let us consider whether the mathematical framework presented earlier provides any basis for anticipating such an interaction. Consider first the case where  $p$  is fixed and  $r$  varies. The potential role of sample size in this context can be seen by considering the structure in Equation 13 and recalling that a major source of sampling error is the presence of non-zero correlations in  $C_{cu}$ ,  $C_{uc}$ , and  $C_{uu}$ . For constant  $p$ , the matrix being fit,  $C_{yy}$ , remains the same size, but as  $r$  becomes smaller, matrices  $C_{cu}$  and  $C_{uc}$  become smaller and thus yield reduced sampling error. Note also that as  $r$  is reduced, the number of parameters decreases, and the impact of sampling error is reduced when fewer parameters are to be estimated. This perspective provides an explicit theoretical basis for the prediction that the effect of sampling error will be

reduced as the number of factors is reduced, holding the number of variables constant.

Next consider the case where  $r$  is fixed and  $p$  increases, thus increasing the  $p:r$  ratio. Although it has been argued in the literature that a larger  $p:r$  ratio is always desirable (e.g., Comrey & Lee, 1992; Bernstein & Nunnally, 1994), present developments may call that view into question. Although a larger  $p:r$  ratio implies potentially better overdetermination of the factors, such an increase also increases the size of the matrices  $C_{cu}$ ,  $C_{uc}$ , and  $C_{uu}$  and in turn increases their contribution to overall sampling error. However, it is also true that as  $p$  increases, the matrix  $C_{yy}$  to which the model is being fit increases in size, thus possibly attenuating the effects of sampling error. In general, though, the nature of the interaction between  $N$  and the  $p:r$  ratio when  $r$  is fixed is not easily anticipated. In their recent Monte Carlo study, Velicer and Fava (1998) held  $r$  constant and varied  $p$ . They found a modest positive effect of increasing  $p$  on recovery of population factor loadings, but they found no interaction with  $N$ .

Finally, our theoretical framework implies a likely interaction between levels of overdetermination and communality. Having highly overdetermined factors might be especially helpful when communalities are low, because it is in that circumstance that the impact of sample size is greatest. Also, when communalities are high, the contents, and thus the size, of the matrices  $C_{cu}$ ,  $C_{uc}$ , and  $C_{uu}$  become almost irrelevant because they receive very low weight in Equation 13. Thus, the effect of the  $p:r$  ratio on quality of factor analysis solutions might well be reduced as communality increases.

### *Summary of Hypotheses*

The theoretical framework presented here provides a basis for the following hypotheses about effects of sample size in factor analysis.

1. As  $N$  increases, sampling error will be reduced, and sample factor analysis solutions will be more stable and will more accurately recover the true population structure.
2. Quality of factor analysis solutions will improve as communalities increase. In addition, as communalities increase, the influence of sample size on quality of solutions will decline. When communalities are all high, sample size will have relatively little impact on quality of solutions, meaning that accurate recovery of population solutions may be obtained using a fairly small sample. However, when communalities are low,

the role of sample size becomes much more important and will have a greater impact on quality of solutions.

3. Quality of factor analysis solutions will improve as overdetermination of factors improves. This effect will be reduced as communalities increase and may also interact with sample size.

### Monte Carlo Study of Sample Size Effects

This section presents the design and results of a Monte Carlo study that investigated the phenomena discussed in the previous sections. The general approach used in this study involved the following steps: (a) Population correlation matrices were defined as having specific desired properties and known factor structures; (b) sample correlation matrices were generated from those populations, using various levels of  $N$ ; (c) the sample correlation matrices were factor analyzed; and (d) the sample factor solutions were evaluated to determine how various aspects of those solutions were affected by sample size and other properties of the data. Each of these steps is now discussed in more detail.

#### *Method*

We obtained population correlation matrices that varied with respect to relevant aspects discussed in the previous section, including level of communality and degree of overdetermination of common factors. Some of the desired matrices were available from a large scale Monte Carlo study conducted by Tucker et al. (1969). Population correlation matrices used by Tucker et al. are ideally suited for present purposes. These population correlation matrices were generated under the assumption that the common factor model holds exactly in the population. Number of measured variables, number of common factors, and level of communality were controlled. True factor-loading patterns underlying these matrices were constructed to exhibit clear simple structure, with an approximately equal number of high loadings on each factor; these patterns were realistic in the sense that high loadings were not all equal and low loadings varied around zero. True common factors were specified as orthogonal. Six population correlation matrices were selected for use in the present study. All were based on 20 measured variables but varied in number of factors and level of communality. Number of factors was either three or seven. The three-factor condition represents the case of highly overdetermined factors, and the seven-factor condition represents the case of rela-

tively weakly overdetermined factors. Three levels of communality were used: high, in which communalities took on values of .6, .7, and .8; wide, in which communalities varied over the range of .2 to .8 in .1 increments; and low, in which communalities took on values of .2, .3, and .4. Within each of these conditions, an approximately equal number of variables was assigned each true communality value. These three levels can be viewed as representing three levels of importance of unique factors. High communalities imply low unique variance and vice versa. One population correlation matrix was selected to represent each condition defined by two levels of number of factors and three levels of communality. The six matrices themselves were obtained from a technical report by Tucker, Koopman, and Linn (1967). For full details about the procedure for generating the matrices, the reader is referred to Tucker et al. (1969). The matrices, along with population factor patterns and communalities, can be found at the worldwide web site of Robert C. MacCallum (<http://quantrm2.psy.ohio-state.edu/maccallum>).

To investigate the effect of overdetermination as discussed in the previous section, it was necessary to construct some additional population correlation matrices. That is, because Tucker et al. (1969) held  $p$  constant at 20 and constructed matrices based on different levels of  $r$  (three or seven), it was necessary for us to construct additional matrices to investigate the effect of varying  $p$  with  $r$  held constant. To accomplish this, we used the procedure developed by Tucker et al. (1969) to generate three population correlation matrices characterized by 10 variables and 3 factors. The three matrices varied in level of communality (high, wide, low), as described above. Inclusion of these three new matrices allowed us to study the effects of varying  $p$  with  $r$  held constant, by comparing results based on 10 variables and 3 factors with those based on 20 variables and 3 factors. In summary, a total of nine population correlation matrices were used—six borrowed from Tucker et al. (1967) and three more constructed by using identical procedures. These nine matrices represented three conditions of overdetermination (as defined by  $p$  and  $r$ ), with three levels of communality.

Sample correlation matrices were generated from each of these nine population correlation matrices, using a procedure suggested by Wijsman (1959). Population distributions were defined as multivariate normal. Sample size varied over four levels: 60, 100, 200, and 400. These values were chosen on the basis



of pilot studies that indicated that important phenomena were revealed using this range of  $N$ s. For data sets based on  $p = 20$  variables, these  $N$ s correspond to  $N:p$  ratios of 3.0, 5.0, 10.0, and 20.0, respectively; for data sets based on  $p = 10$  variables, the corresponding ratios are 6.0, 10.0, 20.0, and 40.0. According to the recommendations discussed in the first section of this article, these  $N$ s and  $N:p$  ratios represent values that range from those generally considered insufficient to those considered more than acceptable. For each level of  $N$  and each population correlation matrix, 100 sample correlation matrices were generated. Thus, with 100 samples drawn from each condition defined by 3 levels of communality, 3 levels of overdetermination, and 4 levels of sample size, a total of 3,600 sample correlation matrices were generated.

Each of these sample matrices was then analyzed using maximum likelihood factor analysis and specifying the number of factors retained as equal to the known number of factors in the population (i.e., either three or seven).<sup>3</sup> No attempt was made to determine the degree to which the decision about the number of factors would be affected by the issues under investigation. Although this is an interesting question, it was considered beyond the scope of the present study.

An issue of concern in the Monte Carlo study involved how to deal with samples wherein the maximum likelihood factor analysis did not converge or yielded negative estimates of one or more unique variances (Heywood cases). Such solutions were expected under some of the conditions in our design. Several studies of confirmatory factor analysis and structural equation modeling (Anderson & Gerbing, 1984; Boomsma, 1985; Gerbing & Anderson, 1987; Velicer & Fava, 1998) have found nonconvergence and improper solutions to be more frequent with small  $N$  and poorly defined latent variables. McDonald (1985, pp. 78–81) suggested that Heywood cases arise more frequently when overdetermination is poor. Marsh and Hau (in press) discussed strategies for avoiding such problems in practice. In the present study, we managed this issue by running the entire Monte Carlo study twice. In the first approach, the *screened-samples approach*, sample solutions that did not converge by 50 iterations or that yielded Heywood cases were removed from further analysis. The sampling procedure was repeated until 100 matrices were obtained for each cell of the design, such that subsequent analyses yielded convergent solutions and no Heywood cases. In the second approach, the *unscreened-samples approach*, the first 100 samples generated in

each cell were used regardless of convergence problems or Heywood cases.

One objective of our study was to compare the solution obtained from each sample correlation matrix with the solution obtained from the corresponding population correlation matrix. To carry out such a comparison, it was necessary to consider the issue of rotation, because all sample and population solutions could be freely rotated. Each population correlation matrix was analyzed by maximum likelihood factor analysis, and the  $r$ -factor solution was rotated using direct quartimin rotation (Jennrich & Sampson, 1966). This oblique analytical rotation was used to simulate good practice in empirical studies. Although population factors were orthogonal in Tucker et al.'s (1969) design, the relationships among the factors would be unknown in practice, thus making a less restrictive oblique rotation more appropriate. Next, population solutions were used as target matrices, and each sample solution was subjected to oblique target rotation. The computer program TARROT (Browne, 1991) was used to carry out the target rotations.

For each of these rotated sample solutions, a measure was calculated to assess the correspondence between the sample solution and the corresponding population solution. This measure was obtained by first calculating a coefficient of congruence between each factor from the sample solution and the corresponding factor from the population solution. If we define  $f_{jk(t)}$  as the true population factor loading for variable  $j$  on factor  $k$  and  $f_{jk(s)}$  as the corresponding sample loading, the coefficient of congruence between factor  $k$  in the sample and the population is given by

---

<sup>3</sup> We chose the maximum likelihood method of factoring because the principles underlying that method are most consistent with the focus of our study. Maximum likelihood estimation is based on the assumption that the common factor model holds exactly in the population and that the measured variables follow a multivariate normal distribution in the population, conditions that are inherent in the simulation design and that imply that all lack of fit and error of estimation are due to sampling error, which is our focus. However, even given this perspective, we know of no reason to believe that the pattern of our results would differ if we had used a different factoring method, such as principal factors. To our knowledge, there is nothing in the theoretical framework underlying our hypotheses that would be dependent on the method of factoring.

$$\phi_k = \frac{\sum_{j=1}^p f_{jk(s)} f_{jk(t)}}{\sqrt{\left(\sum_{j=1}^p f_{jk(s)}^2\right) \left(\sum_{j=1}^p f_{jk(t)}^2\right)}} \quad (15)$$

In geometric terms, this coefficient is the cosine of the angle between the sample and population factors when plotted in the same space. To assess degree of congruence across  $r$  factors in a given solution, we computed the mean value of  $\phi_k$  across the factors. This value will be designated  $K$ :

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

Higher values of  $K$  are indicative of closer correspondence between sample and population factors or, in other words, more accurate recovery of the population factors in the sample solution. Several authors (e.g., Gorsuch, 1983; Mulaik, 1972) have suggested a rule of thumb whereby good matching of factors is indicated by a congruence coefficient of .90 or greater. However, on the basis of results obtained by Korth and Tucker (1975), as well as accumulated experience and information about this coefficient, we believe that a more fine-grained and conservative view of this scale is appropriate. We follow guidelines suggested by Ledyard Tucker (personal communication, 1987) to interpret 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*.

The second aim of our study was to assess variability of factor solutions over repeated sampling within each condition. Within each cell of the design, a mean rotated solution was obtained for the 100 samples. Let this solution be designated  $\bar{B}$ . For a given sample, a matrix of differences was calculated between the loadings in the sample solution  $B$  and the mean loadings in  $\bar{B}$ . The root mean square of these loadings was then computed. We designate this index as  $V$  and define it as follows:

$$V = \left[ \frac{\text{Trace} [(B - \bar{B})' (B - \bar{B})]}{pr} \right]^{1/2} \quad (17)$$

As the sample solutions in a given condition become more unstable, the values of  $V$  will increase for those samples.

To summarize the design of the Monte Carlo study, 100 data sets were generated under each of 36 con-

ditions defined by (a) three levels of population communality (high, wide, low), (b) three levels of overdetermination of the factors ( $p:r = 10:3, 20:3, 20:7$ ), and (c) four levels of sample size (60, 100, 200, 400). The resulting 3,600 samples were then each analyzed by maximum likelihood factor analysis with the known correct number of factors specified. Each of the resulting 3,600 solutions was then rotated to the corresponding direct-quartimin population solution, using oblique target rotation. For each of the rotated sample solutions, measures  $K$  and  $V$  were obtained to assess recovery of population factors and stability of sample solutions, respectively. The entire study was run twice, once screening for nonconvergence and Heywood cases and again without screening.

*Results*

For the screened-samples approach to the Monte Carlo study, Table 1 shows the percentage of factor analyses that yielded convergent solutions and no Heywood cases for each cell of the design. Under the most detrimental conditions ( $N = 60, p:r = 20:7$ , low communality) only 4.1% of samples yielded convergent solutions with no Heywood cases. The percentages increase more or less as  $N$  increases, as communalities increase, and as the  $p:r$  ratio increases. These results are consistent with findings by Velicer and Fava (1998) and others. The primary exception to these trends is that percentages under the 20:7  $p:r$  ratio are much lower than under the 10:3  $p:r$  ratio, even though these ratios are approximately equal. The important point here is that nonconvergence and Heywood cases can occur with high frequency as a result

Table 1  
*Percentages of Convergent and Admissible Solutions in Our Monte Carlo Study*

Ratio of variables to factors and communality level	Sample size			
	60	100	200	400
10:3 ratio				
Low communality	74.6	78.7	95.2	99.0
Wide communality	99.0	98.0	99.0	98.0
High communality	100	100	100	100
20:3 ratio				
Low communality	87.0	97.1	100	100
Wide communality	100	100	100	100
High communality	100	100	100	100
20:7 ratio				
Low communality	4.1	15.8	45.9	80.7
Wide communality	16.5	42.9	72.5	81.3
High communality	39.7	74.6	91.7	97.1

of only sampling error, especially when communalities are low and  $p$  and  $r$  are high. In addition, the frequency of Heywood cases and nonconvergence is exacerbated when poor overdetermination of factors is accompanied by small  $N$ .

We turn now to results for the two outcomes of interest, the congruence measure  $K$  and the variability measure  $V$ . We found that all trends in results were almost completely unaffected by screening in the sampling procedure (i.e., by elimination of samples that yielded nonconvergent solutions or Heywood cases). Mean trends and within-cell variances were highly similar for both screened and unscreened approaches. For example, in the 36 cells of the design, the difference between the mean value of  $K$  under the two methods was never greater than .02 and was less than .01 in 31 cells. Only under conditions of low communality and poor overdetermination, where the proportion of nonconvergent solutions and Heywood cases was highest, did this difference exceed .01, with recovery being slightly better under screening of samples. The same pattern of results was obtained for  $V$ , as well as for measures of within-cell variability of both  $K$  and  $V$ . In addition, effect-size measures in analyses of variance (ANOVAs) of these two indexes never differed by more than 1% for any effect. Thus, results were only trivially affected by screening for nonconvergent solutions and Heywood cases. We choose to present results for the screened-samples Monte Carlo study.

To evaluate the effects discussed in the *Implications for Effects of Sample Size* section, the measures  $K$  and  $V$  were treated as dependent variables in ANOVAs. Independent variables in each analysis were (a) level of communality, (b) overdetermination condition, and (c) sample size. Table 2 presents results of the three-way ANOVA for  $K$ , including effect-size estimates using  $\hat{\omega}^2$  values. This measure pro-

vides an estimate of the proportion of variance accounted for in the population by each effect (Maxwell & Delaney, 1990). Note that all main effects and interactions were highly significant. The largest effect was the main effect of level of communality ( $\hat{\omega}^2 = .41$ ), and substantial effect-size estimates also were obtained for the other two main effects—.11 for the overdetermination condition and .15 for sample size. For the interactions, some of the effect sizes were also substantial—.05 for the Overdetermination  $\times$  Communality interaction and .08 for the Communality  $\times$  Sample Size interaction. The effect sizes for the interaction between overdetermination and sample size and for the three-way interaction were small, but they were statistically significant.

Cell means for  $K$  are presented in Figure 1. The error bars around each point represent 95% confidence intervals for each cell mean. The narrowness of these intervals indicates high stability of mean trends represented by these results. Although substantial interactions were found, it is useful to consider the nature of the main effects because they are consistent with our hypotheses and with earlier findings discussed above. Congruence improved with larger sample size ( $\hat{\omega}^2 = .15$ ) and higher level of communality ( $\hat{\omega}^2 = .41$ ). In addition, for any given levels of  $N$  and communality, congruence declined as the  $p:r$  ratio moved from 20:3 to 10:3 to 20:7. This confirms our hypothesis that for constant  $p$  (here 20), solutions will be better for smaller  $r$  (here 3 vs. 7). Also, for fixed  $r$  (here 3), better recovery was found for larger  $p$  (here 20 vs. 10), although this effect was negligible unless communalities were low. This pattern of results regarding the influence of the  $p:r$  ratio cannot be explained simply in terms of the number of parameters, where poorer estimates are obtained as more parameters are estimated. Although poorest results are obtained in the condition where the largest number of

Table 2  
Analysis of Variance Results for the Measure of Congruence ( $K$ ) in Our Monte Carlo Study

Source	df	MS	F	Prob.	$\hat{\omega}^2$
Sample size ( $N$ )	3	0.72	971.25	<.0001	.15
Communality ( $h$ )	2	2.86	3,862.69	<.0001	.41
Overdetermination ( $d$ )	2	0.75	1,009.29	<.0001	.11
$N \times h$	6	0.18	247.12	<.0001	.08
$N \times d$	6	0.03	34.11	<.0001	.01
$h \times d$	4	0.19	260.77	<.0001	.05
$N \times h \times d$	12	0.01	8.04	<.001	.00
Error	3,564	0.001			

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

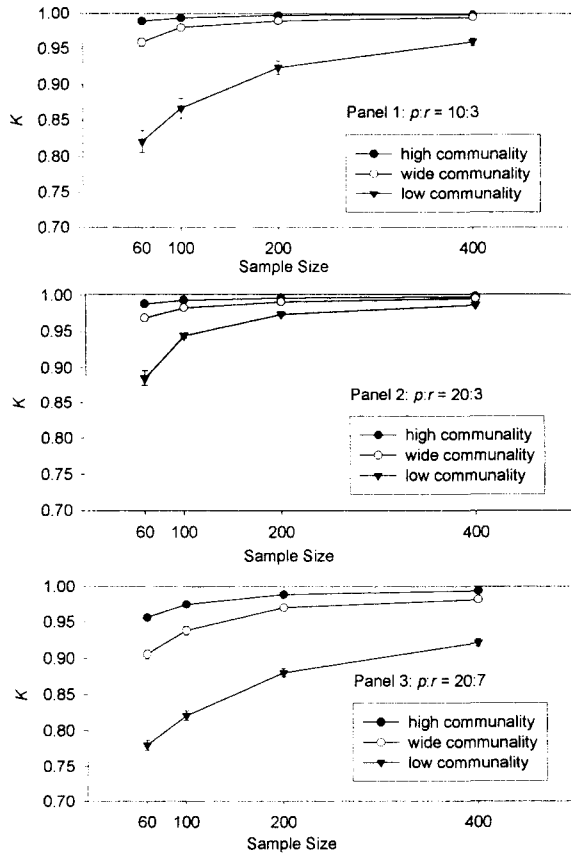


Figure 1 Three-way interaction effect on measure of congruence ( $K$ ) in our Monte Carlo study. Each panel represents a different  $p:r$  ratio. The vertical axis shows mean coefficient of congruence ( $K$ ) between sample and population factors. Error bars show 95% confidence intervals for each cell mean (not visible when width of error bar is smaller than symbol representing mean).  $p$  = the number of variables;  $r$  = the number of factors.

parameters is estimated (when  $p:r = 20:7$ ), better results were obtained when  $p:r = 20:3$  than when  $p:r = 10:3$  even though more parameters are estimated in the former condition.

For the interactive effects, the results in Table 1 and Figure 1 confirm the hypothesized two-way interaction of sample size and level of communality ( $\hat{\omega}^2 = .08$ ). At each  $p:r$  ratio, the effect of  $N$  was negligible for the high communality level, increased somewhat as the communality level became wide, and increased dramatically when the communality level became low. A substantial interaction of levels of communality and overdetermination was also found ( $\hat{\omega}^2 = .05$ ). Figure 1 shows that the impact of communality level on recovery increased as factors became more poorly

determined (i.e., as  $p:r$  moved from 20:3 to 10:3 to 20:7). The interaction of sample size and level of overdetermination was small ( $\hat{\omega}^2 = .01$ ), and the three-way interaction explained less than 0.5% of the variance in  $K$ .

Overall, Figure 1 shows the critical role of communality level. Sample size and level of overdetermination had little effect on recovery of population factors when communalities were high. Recovery of population factors was always good to excellent as long as communalities were high. Only when some or all of the communalities became low did sample size and overdetermination become important determinants of recovery of population factors. On the basis of these findings, it is clear that the optimal condition for obtaining sample factors that are highly congruent with population factors is to have high communalities and strongly determined factors (here,  $p:r = 10:3$  or  $20:3$ ). Under those conditions, 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.

To examine this issue further, recall the guidelines for interpreting coefficients of congruence suggested earlier (L. Tucker, personal communication, 1987). Values of .92 or greater are considered to reflect good to excellent matching. Using this guideline and examining Figure 1, one can see that good recovery of population factors was achieved with a sample size of 60 only when certain conditions were satisfied; specifically, a  $p:r$  ratio of 10:3 or 20:3 and a high or wide range of communalities were needed. A sample size of 100 was adequate at all levels of the overdetermination condition when the level of communalities was either high or wide. However, if the information given in Table 1 is considered, a sample size of 100 may not be large enough to avoid nonconvergent solutions or Heywood cases under a  $p:r$  ratio of 20:7 and a low or wide range of communalities. Finally, a sample size of 200 or 400 was adequate under all conditions except a  $p:r$  ratio of 20:7 and low communalities. These results suggest conditions under which various levels of  $N$  may be adequate to achieve good recovery of population factors in the sample.

Results for the analysis of the other dependent variable, the index of variability, followed the same pattern as results for the analysis of the index of congruence. The most notable difference was that sample

size had a larger main effect on  $V$  than on  $K$ , but the nature of the main effects and interactions was essentially the same for both indexes, with solutions becoming more stable as congruence between sample and population factors improved. Given this strong consistency, we do not present detailed results of the analysis of  $V$ .

### Discussion

Our theoretical framework and results show clearly that common rules of thumb regarding sample size in factor analysis are not valid or useful. The minimum level of  $N$ , or the minimum  $N:p$  ratio, needed to assure good recovery of population factors is not constant across studies but rather is dependent on some aspects of the variables and design in a given study. Most importantly, level of communality plays a critical role. When communalities are consistently high (probably all greater than .6), then that aspect of sampling that has a detrimental effect on model fit and precision of parameter estimates receives a low weight (see Equation 13), thus greatly reducing the impact of sample size and other aspects of design. Under such conditions, recovery of population factors can be very good under a range of levels of overdetermination and sample size. Good recovery of population factors can be achieved with samples that would traditionally be considered too small for factor analytic studies, even when  $N$  is well below 100. Note, however, that with such small samples, the likelihood of nonconvergent or improper solutions may increase greatly, depending on levels of communality and overdetermination (see Table 1). Investigators must not take our findings to imply that high-quality factor analysis solutions can be obtained routinely using small samples. Rather, communalities must be high, factors must be well determined, and computations must converge to a proper solution.

As communalities become lower, the roles of sample size and overdetermination become more important. With communalities in the range of .5, it is still not difficult to achieve good recovery of population factors, but one must have well-determined factors (not a large number of factors with only a few indicators each) and possibly a somewhat larger sample, in the range of 100 to 200. When communalities are consistently low, with many or all under .5, but there is high overdetermination of factors (e.g., six or seven indicators per factor and a rather small number of factors), one can still achieve good recovery

of population factors, but larger samples are required—probably well over 100. With low communalities, a small number of factors, and just three or four indicators for each factor, a much larger sample is needed—probably at least 300. Finally, under the worst conditions of low communalities and a larger number of weakly determined factors, any possibility of good recovery of population factors probably requires very large samples, well over 500.

This last observation may seem discouraging to some practitioners of exploratory factor analysis. It serves as a clear caution against factor analytic studies of large batteries of variables, possibly with low communalities, and the retention of large numbers of factors, unless an extremely large sample is available. One general example of such a study is the analysis of large and diverse batteries of test items, which tend to have relatively low communalities. Under such conditions, the researcher can have very little confidence in the quality of a factor analytic solution unless  $N$  is very large. Such designs and conditions should be avoided in practice. Rather, researchers should make efforts to reduce the number of variables and number of factors and to assure moderate to high levels of communality.

The general implication of our theoretical framework and results is that researchers should carefully consider these aspects during the design of a study and should use them to help determine the size of a sample. In many factor analysis studies, the investigator has some prior knowledge, based on previous research, about the level of communality of the variables and the number of factors existing in the domain of study. This knowledge can be used to guide the selection of variables to obtain a battery characterized by as high a level of communality as possible and a desirable level of overdetermination of the factors. On the basis of our findings, it is desirable for the mean level of communality to be at least .7, preferably higher, and for communalities not to vary over a wide range. For overdetermination, it appears best to avoid situations where both the number of variables and the number of factors are large, especially if communalities are not uniformly high. Rather, it is preferable to analyze smaller batteries of variables with moderate numbers of factors, each represented by a number of valid and reliable indicators. Within the range of indicators studied (three to seven per factor), it is better to have more indicators than fewer. This principle is consistent with the concept of content validity, wherein it is desirable to have larger numbers of variables

(or items) to adequately represent the domain of each latent variable. The critical point, however, is that these indicators must be reasonably valid and reliable. Just as content validity may be harmed by introducing additional items that are not representative of the domain of interest, so recovery of population factors will be harmed by the addition of new indicators with low communalities. Thus, increasing the number of indicators per factor is beneficial only to the extent that those new indicators are good measures of the factors.

Of course, in the early stages of factor analytic research in a given domain, an investigator may not be able to even guess at the level of communality of the variables or the number of factors present in a given battery, thus making it impossible to use any of the information developed in this or future similar studies on an a priori basis. In such a case, we recommend that the researcher obtain as large a sample as possible and carry out the factor analysis. Researchers should avoid arbitrary expansion of the number of variables. On the basis of the level of communalities and overdetermination represented by the solution, one could then make a post hoc judgment of the adequacy of the sample size that was used. This information would be useful in the evaluation of the solution and the design of future studies. If results show a relatively small number of factors and moderate to high communalities, then the investigator can be confident that obtained factors represent a close match to population factors, even with moderate to small sample sizes. However, if results show a large number of factors and low communalities of variables, then the investigator can have little confidence that the resulting factors correspond closely to population factors unless sample size is extremely large. Efforts must then be made to reduce the battery of variables, retaining those that show evidence of being the most valid and reliable indicators of underlying factors.

We believe that our theoretical framework and Monte Carlo results help to explain the widely discrepant findings and rules of thumb about sample size, which we discussed early in this article. Earlier researchers considering results obtained from factor analysis of data characterized by different levels of communality,  $p$ , and  $r$  could arrive at highly discrepant views about minimum levels of  $N$ . For example, it seems apparent that the findings of Barrett and Kline (1981) and Arrindell and van der Ende (1985) of high-quality solutions in small samples must have been based on the analysis of data with relatively high com-

munalities and a desirable level of overdetermination of the factors. Had they used data in which communalities were low and the factors were poorly determined, they undoubtedly would have reached quite different conclusions.

Finally, our theoretical framework and the results of our Monte Carlo study provide validation, as well as a clear explanation, for results obtained previously by other investigators. Most importantly, our framework accounts for influences of communality on quality of factor analysis solutions observed by Browne (1968), Cliff and Pennell (1967), Velicer and Fava (1987, 1998), and others, as well as the interactive effect of sample size and communality level observed by Velicer and Fava (1987, 1998). Although the validity of such findings had not been questioned, a clear theoretical explanation for them had not been available. In addition, some authors had previously associated improved recovery of factors with higher loadings (Velicer & Fava, 1987, 1998). Cliff and Pennell had showed experimentally that improvement was due to communality size rather than loading size per se, and our approach verifies their argument. Our approach also provides a somewhat more complete picture of the influence of overdetermination on recovery of factors. Such effects had been observed previously (Browne, 1968; Tucker et al., 1969; Velicer & Fava, 1987, 1998). We have replicated these effects and have also shown interactive effects of overdetermination with communality level and sample size.

### Limitations and Other Issues

The theoretical framework on which this study is based focuses only on influences of sampling error, assuming no model error. The assumption of no model error is unrealistic in practice, because no parsimonious model will hold exactly in real populations. The present study should be followed by an investigation of the impact of model error on the effects described herein. Such a study could involve analysis of empirical data as well as artificial data constructed to include controlled amounts of model error. We have conducted a sampling study with empirical data that has yielded essentially the same pattern of effects of sample size and communality found in our Monte Carlo study. Results are not included in this article.

Further study could also be directed toward the interplay between the issues studied here and the number-of-factors question. In our Monte Carlo

study, we retained the known correct number of factors in the analysis of each sample. With this approach, it was found that excellent recovery of population factors could be achieved with small samples under conditions of high communality and optimal overdetermination of factors. However, it is an open question whether analysis of sample data under such conditions would consistently yield a correct decision about the number of factors. We expect that this would, in fact, be the case simply because it would seem contradictory to find the number of factors to be highly ambiguous but recovery of population factors to be very good if the correct number were retained. Nevertheless, if this were not the case, our results regarding recovery of population factors under such conditions might be overly optimistic because of potential difficulty in identifying the proper number of factors to retain.

One other limitation of our Monte Carlo design involves the limited range of levels of  $N$ ,  $p$ , and  $r$  that were investigated. Strictly speaking, one must be cautious in extrapolating our results beyond the ranges studied. However, it must be recognized that our findings were supported by a formal theoretical framework that was not subject to a restricted range of these aspects of design, thus lending credence to the notion that the observed trends are probably valid beyond the conditions investigated in the Monte Carlo study.

Our approach to studying the sample size question in factor analysis has focused on the particular objective of obtaining solutions that are adequately stable and congruent with population factors. A sample size that is sufficient to achieve that objective might be somewhat different from one that is sufficient to achieve some other objective, such as a specified level of power for a test of model fit (MacCallum, Browne, & Sugawara, 1996) or standard errors of factor loadings that are adequately small by some definition. An interesting topic for future study would be the consistency of sample sizes required to meet these various objectives.

We end with some comments about exploratory versus confirmatory factor analysis. Although the factor analysis methods used in our simulation study were exploratory, the issues and phenomena studied in this article are also relevant to confirmatory factor analysis. Both approaches use the same factor analysis model considered in our theoretical framework, and we expect that sampling error should influence solutions in confirmatory factor analysis in much the same fashion as observed in the exploratory factor

analysis setting. One relevant distinction between typical exploratory and confirmatory studies is that the latter might often be characterized by indicators with higher communalities, because indicators in such studies are often selected on the basis of established quality as measures of constructs of interest. As a result, influences of sampling error described in this article might be attenuated in confirmatory studies.

## References

- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155–173.
- Archer, J. O., & Jennrich, R. I. (1973). Standard errors for orthogonally rotated factor loadings. *Psychometrika*, *38*, 581–592.
- Archer, J. O., & Jennrich, R. I. (1976). A look, by simulation, at the validity of some asymptotic distribution results for rotated loadings. *Psychometrika*, *41*, 537–541.
- 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.
- Bernstein, I. H., & Nunnally, J. C. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, *50*, 229–242.
- Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika*, *33*, 267–334.
- Browne, M. W. (1991). TARROT: A computer program for carrying out an orthogonal or oblique rotation to a partially specified target [Computer software]. Columbus, OH: Author.
- Browne, M. W., & Cudeck, R. (1997, July). *A general approach to standard errors for rotated factor loadings*. Paper presented at the European Meeting of the Psychometric Society, Santiago de Compostelo, Spain.
- Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum.
- Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16 PF)*. Champaign, IL: Institute for Personality and Ability Testing.
- Cliff, N., & Hamburger, C. D. (1967). The study of sam-

- pling errors in factor analysis by means of artificial experiments. *Psychological Bulletin*, 68, 430–445.
- Cliff, N., & Pennell, R. (1967). The influence of communality, factor strength, and loading size on the sampling characteristics of factor loadings. *Psychometrika*, 32, 309–326.
- 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.
- Cudeck, R., & O’Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin*, 115, 475–487.
- Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry*, 126, 237–240.
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Eysenck Personality Questionnaire manual*. San Diego, CA: Educational and Industrial Testing Service.
- Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecification. *Psychometrika*, 52, 99–111.
- Girshick, M. A. (1939). On the sampling theory or roots of determinantal equations. *Annals of Mathematical Statistics*, 10, 203–224.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.
- Jennrich, R. I. (1973). Standard errors for obliquely rotated factor loadings. *Psychometrika*, 38, 593–604.
- Jennrich, R. I. (1974). Simplified formulae for standard errors in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, 27, 122–131.
- Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, 31, 313–323.
- Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.
- Korth, B., & Tucker, L. R. (1975). The distribution of chance congruence coefficients from simulated data. *Psychometrika*, 40, 361–372.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. New York: Macmillan.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size in covariance structure modeling. *Psychological Methods*, 1, 130–149.
- 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.
- Marsh, H. W., & Hau, K. (in press). Confirmatory factor analysis: Strategies for small sample sizes. In R. Hoyle (Ed.), *Statistical issues for small sample research*. Thousand Oaks, CA: Sage.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data*. Belmont, CA: Wadsworth.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.
- Morrison, D. F. (1990). *Multivariate statistical methods* (3rd ed.). New York: McGraw-Hill.
- Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.
- Pennell, R. (1968). The influence of communality and  $N$  on the sampling distributions of factor loadings. *Psychometrika*, 33, 423–439.
- Tucker, L. R., Koopman, R. F., & Linn, R. L. (1967). *Evaluation of factor analytic research procedures by means of simulated correlation matrices* (Tech. Rep.). Champaign: University of Illinois, Department of Psychology.
- 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. (1987). An evaluation of the effects of variable sampling on component, image, and factor analysis. *Multivariate Behavioral Research*, 22, 193–210.
- 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.
- Wijsman, R. A. (1959). Applications of a certain representation of the Wishart matrix. *Annals of Mathematical Statistics*, 30, 597–601.

Received December 23, 1997

Revision received June 9, 1998

Accepted November 7, 1998 ■