Behavior Genetics, Vol. 32, No. 2, March 2002 (© 2002)

Exploratory Factor Analysis in Behavior Genetics Research: Factor Recovery with Small Sample Sizes

Kristopher J. Preacher[1,2] and Robert C. MacCallum[1]

Received 19 Nov. 2001—Final 28 Jan. 2002

Results of a Monte Carlo study of exploratory factor analysis demonstrate that in studies characterized by low sample sizes the population factor structure can be adequately recovered if communalities are high, model error is low, and few factors are retained. These are conditions likely to be encountered in behavior genetics research involving mean scores obtained from sets of inbred strains. Such studies are often characterized by a large number of measured variables relative to the number of strains used, highly reliable data, and high levels of communality. This combination of characteristics has special consequences for conducting factor analysis and interpreting results. Given that limitations on sample size are often unavoidable, it is recommended that researchers limit the number of expected factors as much as possible.

KEY WORDS: Factor analysis; inbred strains; strain means; sample size; communality; factor recovery.

Researchers in the field of behavior genetics who use inbred strains in breeding designs, or who examine strain distribution patterns of recombinant inbred strain means, often hesitate to employ exploratory factor analysis (EFA) to investigate the number and nature of unobservable genetic factors underlying patterns of correlations among the strain means of their variables. This reluctance is not surprising, given the various "rules of thumb" encountered in the literature regarding the sample size, *N*, required for factor analysis, where *N* in these cases represents the number of strains involved in the study. Popular recommendations regarding the minimum necessary sample size range from $N = 100$ (e.g., Gorsuch, 1983) to $N = 250$ (e.g., Cattell, 1978), values well beyond the number of strains typically employed in behavior genetics research. Similarly, recommendations regarding the proper factor-to-variable ratio range from 3–6 (Cattell, 1978) to at least 10 (Everitt, 1975).

Use of factor analysis can be valuable in behavior genetic research, where a given study will often include many dependent measures but few underlying latent variables. Failure to first perform factor analysis on dependent measures derived from test batteries and multidimensional scales has been shown to adversely affect interpretation of complex behaviors, such as emotionality and intelligence (Royce, Holmes, and Poley, 1975). Whenever a multidimensional construct is treated as unidimensional in statistical analyses, proper interpretation of results is nearly impossible, and solutions are unstable across similar studies. EFA can be used as an intermediate step to help identify clusters of variable correlated because of a mutual dependence on underlying latent variables, which in turn can guide researchers in test battery development, variable selection, and interpretation of results.

The typical potential application of EFA in animal behavior genetics involves those cases where a large number of dependent measures is administered to a limited number of inbred strains or F1 crosses (typically numbering under 30 and sometimes fewer than 10). Whereas these conditions are usually considered far from appropriate for EFA (as we discuss in detail later), these studies often have redeeming characteristics. For example, the "individual scores" submitted to an EFA of genetic factors are group means derived from many individuals, and thus provide relatively accurate (highly

[1] Department of Psychology, The Ohio State University.
[2] To whom correspondence should be addressed: 142 Townshend Hall, 1885 Neil Ave. Mall, The Ohio State University, Columbus, OH 43210-1222. e-mail: preacher.2@osu.edu

0001-8244/02/0300-0153/0 © 2002 Plenum Publishing Corporation

reliable) characterizations of individual strains on the dependent measures.

The goal of EFA is to identify a limited number of underlying (latent) variables responsible for observed variances and covariances. The number of factors to retain is decided upon by the researcher, and is ideally based on a combination of statistical and interpretational criteria (Fabrigar, Wegener, MacCallum, and Strahan, 1999). EFA is not to be confused with principal components analysis (PCA), which refers to a model with a different set of applications and an underlying philosophy distinct from that of EFA (Floyd and Widaman, 1995; Preacher and MacCallum, 2000).

DETERMINANTS OF SUCCESSFUL APPLICATION OF EFA

The utility of EFA hinges on its ability to yield stable, accurate, and interpretable estimates of factor loadings. One of the most common questions that arises in applications of EFA regards the minimum sample size required for the analysis. This question can be examined from the statistical power perspective [i.e., how many cases are required to achieve a given level of power for statistical tests of model fit (MacCallum, Browne, and Sugawara, 1996)], the precision perspective (i.e., how many cases are necessary to achieve acceptably stable parameter estimates), or from the validity perspective (i.e., how many cases are required in order to recover the population factor structure). It is upon the third perspective that this paper focuses, because population factor recovery is arguably the most important requirement for good interpretation and inference.

Several model and design characteristics play a role in EFA's ability to adequately recover population factor loadings. Specifically, the sample size (N) employed, the number of factors (m) retained, the number of variables (p) examined, the level of communality (h), and the degree of model error[3] each contribute to factor recovery (MacCallum, Widaman, Zhang, and Hong, 1999; MacCallum, Widaman, Preacher, and Hong, 2002). Previous studies examining the influence of N have shown that its effects depend on levels of other design characteristics. Therefore, it should be evident that general rules of thumb regarding the minimum required N are not valid. In applications of EFA encountered in behavior genetics research, characteristics of factor models commonly include small Ns, a small factor-to-variable ratio (low overdetermination), and high communalities. This combination of characteristics has consequences for the degree of successful factor recovery researchers can expect to achieve.

Using a framework established by MacCallum and Tucker (1991) to represent the various sources of error in factor analysis, MacCallum *et al.* (1999) used a Monte Carlo approach to investigate the determinants of population factor recovery. They demonstrated that at high levels of communality (in the neighborhood of $h = .6$ to .8) low overdetermination contributes to relatively poor factor recovery and that factor recovery drops off only slightly as N is lowered. At lower levels of h, on the other hand, factor recovery suffers greatly as a function of low N and low overdetermination. High communalities, then, tend to offset the deleterious effects of small sample sizes and low overdetermination. MacCallum *et al.* (2002) followed this research by examining model error (the degree to which a model accurately represents relationships among variables in the population) as a possible determinant of factor recovery. Besides replicating effects observed by MacCallum *et al.* (1999), they found that lack of fit of the common factor model in the population did not seriously influence the degree of factor recovery. In research in behavior genetics, sample sizes are usually quite low, but communalities are typically high (over .8), and overdetermination and model fit are likewise usually high, so there is still a good chance of adequately recovering population factors with small Ns. However, small Ns can have a negative effect on aspects of a factor analysis other than factor recovery.

One consequence of employing small Ns is that researchers often cannot take advantage of model fit indices requiring the chi-square likelihood ratio test statistic, which is obtained through maximum likelihood (ML) parameter estimation. ML estimation involves the iterative minimization of a function involving S^{-1}, the inverse of the sample covariance matrix S. The result of this process is a test statistic distributed as chi-square (χ^2). This χ^2 statistic can be used as a test of model fit for a given number of factors m, and can be used as one criterion to help determine the appropriate number of factors to retain. The matrix S^{-1}, a necessary component of the ML minimization function, is undefined when the number of variables exceeds N. Inability to use ML estimation prohibits the use of χ^2, which in turn prohibits the use of many fit statistics, such as the root

[3] Model error is defined as the lack of fit of the model in the population. Model error is independent of sampling error.

mean square error of approximation (RMSEA; Steiger and Lind, 1980), which are useful in determining not only the degree to which the model is consistent with data, but in deciding how many factors to retain and in conducting power analyses. A second consequence of using small *N*s is the danger of obtaining unstable estimates of factor loadings, an issue still warranting investigation.

MONTE CARLO STUDY

Because small sample sizes relative to the numbers of dependent measures are common characteristics of correlational research using strain means in behavior genetics, there may be concerns about the use of EFA in this area. However, because highly reliable data and high communality levels are also likely, good factor recovery may still be achieved in some situations despite small sample sizes and large test batteries. It is possible that some researchers, fearing that their data are lacking in some respect, avoid using EFA when it would have been both appropriate and informative or that some legitimate uses of EFA with small *N*s have not passed peer review.

With these concerns in mind, we wished to investigate the degree of population factor recovery under various combinations of sample size (*N*), degree of overdetermination, and model error commonly found in behavior genetics research. Model fit (*f*) was defined in terms of the population root mean squared residual (*RMSR*).[4] Overdetermination, the ratio of factors to variables, was separated into number of factors (*m*) and number of variables (*p*) in order to examine the effects of each model characteristic individually.

We seek to extend prior research by examining how *N, f, m,* and *p* influence factor recovery in designs with very small *N*s, a design limitation frequently encountered in behavior genetics research using inbred strains. It is unknown whether the effects observed in prior research will hold under these conditions, but we expect to replicate many of the results from earlier studies (e.g., MacCallum *et. al.,* 1999, 2002). For example, we expect *N* to be a primary determinant of factor recovery. Because overdetermination had a significant effect on factor recovery in prior research, we also expect to see main effects for *m, p,* or both *m* and *p*. We do not expect *f* to have a very large effect (if any), given prior research and the restricted range of model fit we employ.

Method

Communalities were kept uniformly high; *h* ranged between approximately .8 to .9 for each variable in the population. We systematically varied sample size (N = 10, 20, 30, 50), the number of factors retained (m = 2, 4), the number of observed variables (p = 10, 25, 40), and the degree of model fit (f) in the population in terms of the population root mean squared residual (*RMSR* = .00, .03, .06). Respectively, these values of *RMSR* correspond to perfect, good, and fair model fit in the population. The range of *f* was restricted to the "acceptable" region because poor model fit is difficult to reconcile with high communalities. We felt that the chosen ranges for *m, p, f,* and *N* cover the majority of applications of EFA in behavior genetics research.

Using a method developed by Tucker, Koopman, and Linn (1969), 18 population correlation matrices were generated to correspond to the 18 combinations of *m, p,* and *f*. Using a method developed by Wijsman (1959), sample correlation matrices were generated from the population matrices to correspond to each of the four selected *N*s. Enough sample matrices were generated within each of the 72 conditions defined by combinations of *m, p, f,* and *N* so that at least 100 matrices in each cell exhibited no Heywood cases when factor analyzed. All subsequent analyses were conducted on two sets of matrices: one which retained the first 100 matrices showing no Heywood cases (communalities > 1) when factor analyzed (the screened sample) and one retaining the first 100 matrices, regardless of the presence of Heywood cases (the unscreened sample).

Each population matrix was submitted to a factor analysis using the iterative principal factors method (equivalent to ordinary least squares) with oblique direct quartimin rotation (Jennrich and Sampson, 1966), specifying *m* as the number of known population factors. Then all sample matrices were factor analyzed using the iterative principal factors method and retaining *m* factors. The obtained factors were rotated to simple structure by using both direct quartimin and oblique least-squares target rotation, using the population solu-

[4] $RMSR = \sqrt{\frac{2\sum_{i-1}^{p}\sum_{j=1}^{i}(P_{ij} - P_{ij})^2}{p(p+1)}}$, where P is the population correlation matrix and *p* is the number of measured variables. The *RMSR* index yields an estimate of the average degree of discrepancy between corresponding elements of the population correlation matrix and the correlation matrix implied by a factor model with *m* factors.

tion as a target. All factor rotations were conducted using the Comprehensive Exploratory Factor Analysis (CEFA) program of Browne, Cudeck, Tateneni, and Mels (1998). As in MacCallum *et al.* (1999, 2002), ANOVAs were carried out both for samples screened and unscreened for Heywood cases. Because results were virtually identical, we chose to report the results corresponding to the unscreened samples only.

To assess the degree of factor recovery, two indices were computed: *K*, the average coefficient of congruence across the *m* factors (Tucker, 1951); and *g*, the root mean squared deviation between sample and population loadings (Velicer and Fava, 1998). We also computed δ, the mean deviation between corresponding sample and population loadings, as a measure of bias. Details regarding the calculation of these indices can be found in MacCallum *et al.* (2002). Suggested guidelines for the interpretation of *K* (Tucker, personal communication, 1987), are: .98 to 1.00 = *excellent,* .92 to .98 = *good,* .82 to .92 = *borderline,* .68 to .82 = *poor,* and below .68 = *terrible.* In summary, high values of *K* and low values of *g* are considered indicators of good factor recovery, and values of δ that depart from zero indicate the presence of bias.

To summarize the design, 100 sample correlation matrices were generated for each of 72 conditions defined by four levels of sample size, two levels of factors retained, three levels of observed variables, and three levels of model fit. The resulting 7,200 sample matrices were each factor analyzed using the iterative principal factors method, rotated to both oblique direct quartimin and oblique least-squares target solutions, and compared to corresponding population loadings. Measures of sample-population congruence (*K* and *g*) and bias (δ) were obtained.

Results

Indices *K*, *g*, and δ were submitted to a 4 (*N*) × 3 (*p*) × 3 (*f*) × 2 (*m*) ANOVA. As in MacCallum *et al.* (1999, 2002), we expected virtually all effects to be statistically significant, so interpretation is based primarily on effect size as measured by $\hat{\omega}^2$, an estimate of the proportion of variance accounted for in the population by each effect (Maxwell and Delaney, 1990).

As expected, virtually all main effects and interactions in the ANOVAs involving *K*, *g*, and δ were statistically significant. The results of analyses regarding the δ measure of bias indicated a small effect of sample size, such that less bias was associated (not surprisingly) with larger sample sizes. However, we stress that this effect was quite small, so no ANOVA tables or plots are presented for δ. The overall average difference between corresponding sample and population factor loadings was only .012, with the largest average difference (.043) occurring in the design cell corresponding to $p = 40$, $m = 2$, f = fair model fit (*RMSR* = .06), and $N = 10$.

Results of the ANOVAs[5] for the congruence index *K* are reported in Table I.[6] MacCallum *et al.* (1999, 2002) found communality to be the most important determinant of factor recovery. Here, population communalities were held uniformly high. In the present study, sample size was the most important predictor of *K* ($\hat{\omega}^2 = .51$), followed by the number of factors ($\hat{\omega}^2 = .29$), such that samples with larger *N*s and fewer factors tended to show greater degrees of congruence. All other effect size estimates were negligible.

Results of the ANOVAs for the root mean squared deviation index *g* are presented in Table II. The only non-negligible effect was that of sample size ($\hat{\omega}^2 = .66$), such that larger sample sizes demonstrated lower values of *g*.

Cell means for *K* and *g* are presented in Fig. 1 and 2, respectively. Error bars represent lower and upper bounds of asymptotic 95% confidence intervals (e.g., $\bar{K} \pm 1.96(S_{\bar{K}})$).[7] Where comparisons to prior research can be made, these analyses and plots replicate trends demonstrated in MacCallum *et al.* (1999, 2001) in circumstances involving small sample sizes and low levels of model error, conditions commonly encountered in behavior genetics research.

DISCUSSION AND RECOMMENDATIONS

As assessed by both indices *K* and *g*, recovery of the population factor structure was quite good in most conditions examined. The remainder of the discussion is organized by aspects contributing to factor recovery.

---

[5] The same general pattern of results for *K*, *g*, and δ was obtained by using direct quartimin rotated solutions and target rotated solutions. Similarly, the same pattern of results held whether solutions were screened or unscreened for the presence of Heywood cases. Therefore, all ANOVA results and mean plots refer to analyses of target rotated solutions, unscreened for Heywood cases. Results and plots from analyses involving oblique direct quartimin rotated solutions and/or Heywood cases can be found at the authors' website: http://quantrm2.psy.ohio-state.edu/maccallum/pm/results.htm.

[6] Values of ϕ were transformed to normality by means of the Fisher *r*-to-*z* transformation before inclusion of the *K* index in ANOVAs.

[7] These confidence intervals assume normality. *K* and *g*, of course, are not normally distributed, but the sampling distributions of their means are near-normal. Consequently, standard errors of the mean were used to construct intervals. Because the error of estimation is assumed to be symmetric about the mean, only the lower half of each error bar is presented.

Table I. ANOVA Results for the Measure of Congruence (K)

Source	df	MS	F	Prob.	$\hat{\omega}^2$
Model fit (f)	2	1.33	27.36	0.00	0.00
No. of factors (m)	1	566.21	11661.89	0.00	0.29
No. of variables (p)	2	3.58	73.83	0.00	0.00
Sample size (N)	3	330.42	6805.42	0.00	0.51
$f \times m$	2	4.85	99.97	0.00	0.00
$f \times p$	4	1.94	39.95	0.00	0.00
$m \times p$	2	2.27	46.83	0.00	0.00
$f \times m \times p$	4	2.44	50.33	0.00	0.00
$f \times N$	6	0.02	0.48	0.83	0.00
$m \times N$	3	0.25	5.23	0.01	0.00
$f \times m \times N$	6	0.08	1.74	0.11	0.00
$p \times N$	6	0.50	10.26	0.00	0.00
$f \times p \times N$	12	0.10	2.09	0.01	0.00
$m \times p \times N$	6	0.05	1.01	0.41	0.00
$f \times m \times p \times n$	12	0.04	0.92	0.52	0.00
Error	7128	0.05			

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

Model Fit

Consistent with the findings of MacCallum *et al.*, model fit had little effect on factor recovery. This finding is perhaps due to the fact that all three levels of model fit examined (*RMSR* = .00, .03, and .06) represent relatively good fit. As indicated earlier, the range of f was restricted because it is probably very rare in practice to find factor models exhibiting simultaneously high communalities and poor fit, so results involving such a combination would have had limited generalizability.

Sample Size

Within the range of model characteristics included in our design, N had by far the largest effect on factor recovery, which exhibited a sharp drop-off below Ns of 20 or so. To the extent that researchers may collect data from larger samples, they would be well advised to do so. However, we recognize that behavior genetics research involving strain means is often limited in this capacity, and that many designs have Ns so low that they also preclude the use of the χ^2 fit statistic. Given that methods for calculating power have been

Table II. ANOVA Results for the Root Mean Squared Error (g)

Source	df	MS	F	Prob.	$\hat{\omega}^2$
Model fit (f)	2	0.01	14.63	0.00	0.00
No. of factors (m)	1	0.37	686.16	0.00	0.03
No. of variables (p)	2	0.02	32.89	0.00	0.00
Sample size (N)	3	3.03	5600.38	0.00	0.66
$f \times m$	2	0.02	45.23	0.00	0.00
$f \times p$	4	0.01	11.51	0.00	0.00
$m \times p$	2	0.02	37.06	0.00	0.00
$f \times m \times p$	4	0.01	19.93	0.00	0.00
$f \times N$	6	0.00	0.84	0.54	0.00
$m \times N$	3	0.02	45.11	0.00	0.01
$f \times m \times N$	6	0.00	0.94	0.47	0.00
$p \times N$	6	0.01	23.38	0.00	0.01
$f \times p \times N$	12	0.00	1.06	0.39	0.00
$m \times p \times N$	6	0.00	1.49	0.18	0.00
$f \times m \times p \times n$	12	0.00	0.77	0.69	0.00
Error	7128				

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

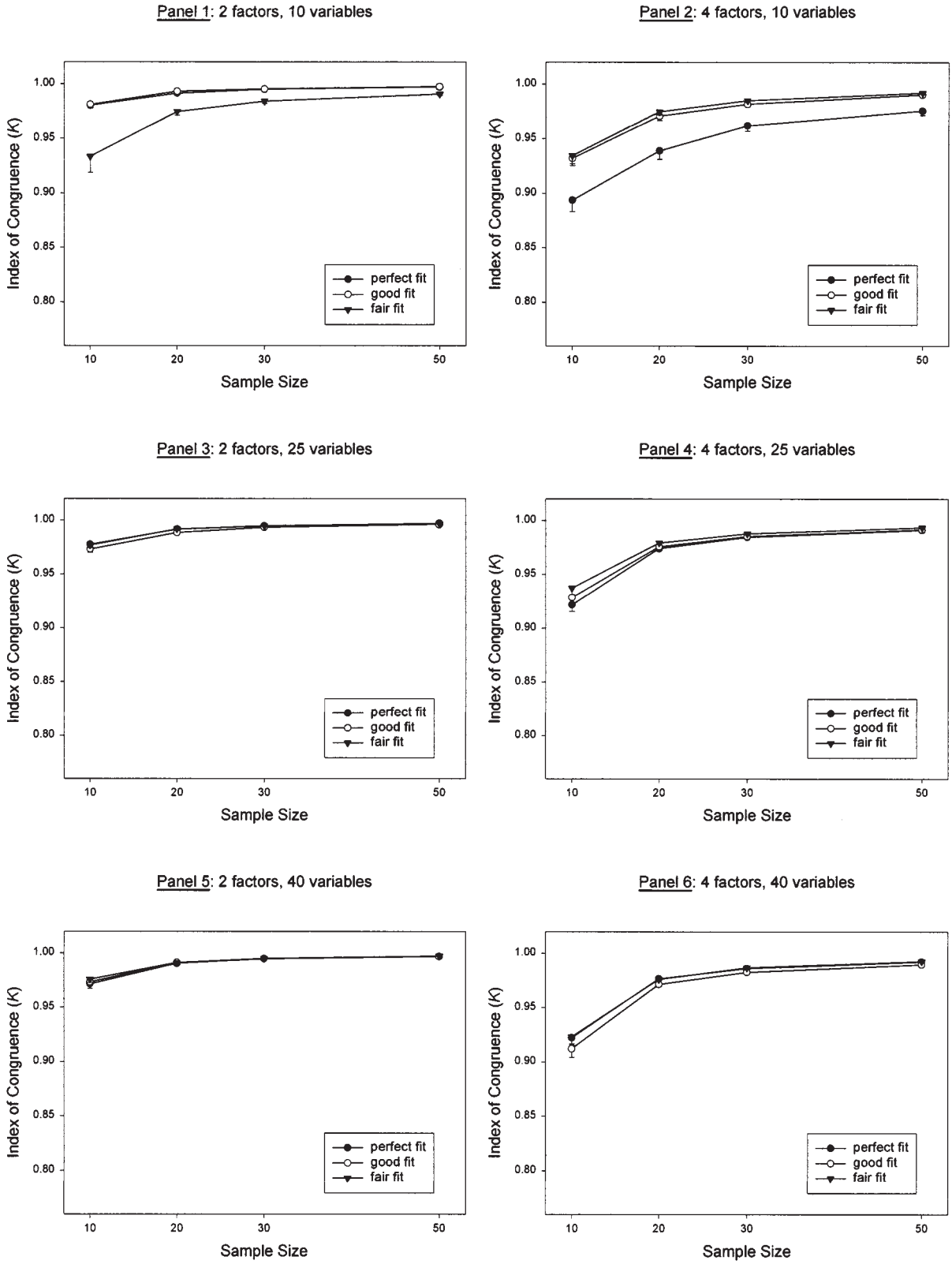

Fig. 1. Cell means for index of congruence (*K*). Each panel represents a different combination of *p*, *m*, *N*, and *f*. The vertical axis shows mean *K* between sample and population factors. Error bars show intervals within which 95% of the values of *K* lie for each cell (not visible when the width of the error bar is smaller than the symbol representing the mean). *p* = number of variables; *m* = number of factors; *N* = sample size; *f* = model fit in terms of *RMSR*.

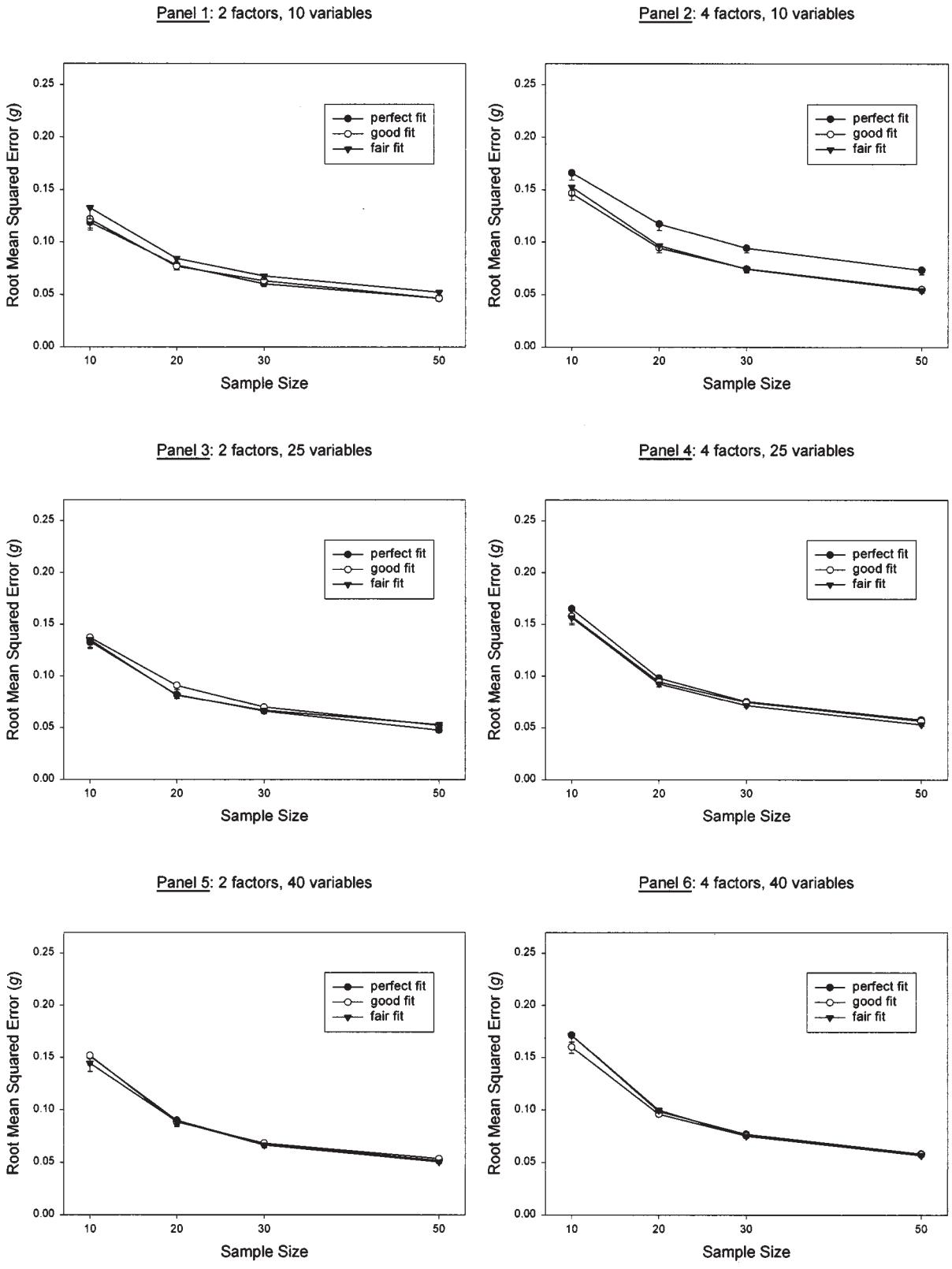

Fig. 2. Cell means for index of root mean squared deviation (*g*).

developed only for those fit statistics involving χ^2 (MacCallum *et al.*, 1996; MacCallum and Hong, 1997), the appropriate *N* must be based on the desired degree of factor recovery, the desired stability of factor loadings, or both. Inspection of Fig. 1 and 2 reveals that good factor recovery may still be achieved despite small *N*s.

Overdetermination

To make comparisons with prior research in this area easier, note that if *p* is held constant, varying *m* results in different levels of factor overdetermination. Thus, holding *p* constant, overdetermination is higher when $m = 2$ than when $m = 4$. The main effect of *m* in the present study thus conceptually replicates the main effect of overdetermination found in earlier studies (MacCallum *et al.*, 1999, 2002), such that better factor recovery is associated with higher overdetermination (fewer factors). The results of the present study suggest that the number of factors, rather than the number of variables, is what drove the overdetermination effect in prior studies, although the role of *p* may be greater in designs with larger samples and/or smaller communalities.

This finding may appear to suggest that researchers are more likely to improve factor recovery by reducing the number of factors rather than by adding indicators, but note that communalities were held constant in the present study. Reducing the number of factors in practical applications will tend to reduce communalities, which may drastically affect factor recovery. Based on the results of MacCallum *et al.* (1999, 2002) we suspect that if communalities were even a little lower, the observed pattern of effects would be exaggerated.

What, then, should be done regarding the number of factors and number of variables? In any application of EFA, retaining too few factors will negatively impact communalities, whereas retaining too many can compromise interpretability. The number of factors explaining interrelationships among variables is often beyond the control of the experimenter and thus not susceptible to manipulation anyway. We recommend that researchers use established methods to determine the number of factors to retain (Fabrigar *et al.*, 1999; Floyd and Widaman, 1995), being careful not to retain factors that add little explanatory power to the model. Our practical finding regarding overdetermination is that, for studies with small *N*s and high communalities such as are commonly found in the field of behavior genetics, designs with a small number of well-determined factors have the best chance of successful factor recovery regardless of the number of variables examined. If factor recovery is poor, adding variables is unlikely to improve matters.

GENERAL RECOMMENDATIONS

Unfortunately for most animal research where the observations of interest are strain means, *N*s are necessarily small. Given that, our recommendation is to design studies characterized by small expected numbers of factors (this is the only reasonable way in which the number of factors may be manipulated). Although only two values of *m* were examined in the present study, factor recovery is clearly better in situations involving fewer underlying factors. Because the number of factors is in some sense beyond the control of the experimenter, it makes sense to design studies in which the number of *expected* factors is small.

The other point we wish to emphasize is that good factor recovery may be achieved even with very small sample sizes, assuming other conditions hold. This conclusion might be somewhat surprising to those familiar with conventional rules of thumb regarding sample size in factor analysis. As long as communalities are high, the number of expected factors is relatively small, and model error is low (a condition which often goes hand-in-hand with high communalities), researchers and reviewers should not be overly concerned about small sample sizes.

ACKNOWLEDGMENTS

We wish to thank Dr. Norman Henderson for clarifying key methodological issues in a field with which we are largely unfamiliar. We also thank Nancy Briggs and Donna Coffman for helpful comments on the manuscript.

REFERENCES

Browne, M. W., Cudeck, R., Tateneni, K., and Mels, G. (1998). CEFA: Comprehensive Exploratory Factor Analysis [Computer software and manual]. Retrieved from http://quantrm2.psy.ohio-state.edu/browne/.

Cattell, R. B. (1978). *The Scientific Use of Factor Analysis*. New York: Plenum.

Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry* **126**:237–240.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* **4**(3):272–299.

Floyd, F. J., and Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* **7**(3):286–299.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
Jennrich, R. I., and Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika* **31**(3):313–323.
MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* **1**:130–149.
MacCallum, R. C., and Hong, S. (1997). Power analysis in covariance structure modeling using GFI and AGFI. *Multivariate Behavioral Research* **32**:193–210.
MacCallum, R. C., and Tucker, L. R (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin* **109**(3):502–511.
MacCallum, R. C., Widaman, K. F., Zhang, S., and Hong, S. (1999). Sample size in factor analysis. *Psychological Methods* **4**: 84–99.
MacCallum, R. C., Widaman, K. F., Preacher, K. J., and Hong, S. (2002). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research* **36**:611–637.
Maxwell, S. E., and Delaney, H. D. (1990). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Pacific Grove, CA: Brooks/Cole Pub. Co.
Preacher, K. J., and MacCallum, R. C. (2000). *Repairing Tom Swift's electric factor analysis machine* [unpublished manuscript].
Royce, J. R., Holmes, T. M., and Poley, W. (1975). Behavior genetic analysis of mouse emotionality. III. The diallel analysis. *Behavior Genetics* **5**(4):351–372.
Steiger, J. H., and 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.
Tucker, L. R (1951). A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984). Washington, DC: Department of the Army.
Tucker, L. R, Koopman, R. F., and Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika* **34**(4):421–459.
Velicer, W. F., and Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods* **3**:231–251.
Wijsman, R. A. (1959). Applications of a certain representation of the Wishart matrix. *Annals of Mathematical Statistics* **30**: 597–601.

Edited by Norman D. Henderson