Structural Equation Modeling, 20:86–97, 2013

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705511.2013.742388

Routledge
Taylor & Francis Group

A Note on Sample Size and Solution Propriety for Confirmatory Factor Analytic Models

Dennis L. Jackson,[1] Jennifer Voth,[1] and Marc P. Frey[1]
[1]*University of Windsor*

Determining an appropriate sample size for use in latent variable modeling techniques has presented ongoing challenges to researchers. In particular, small sample sizes are known to present concerns over sampling error for the variances and covariances on which model estimation is based, as well as for fit indexes and convergence failures. The literature on the topic has focused on conducting power analyses as well as identifying rules of thumb for deciding an appropriate sample size. Often the advice involves an assumption that sample size requirement is moderated by aspects of the model in question. In this study, an effort was undertaken to extend the findings of Gagné and Hancock (2006) on measurement model quality and solution propriety to a broader set of confirmatory factor analysis models. As well, we examined whether Herzog, Boomsma, and Reinecke's (2007) findings for the Swain correction to the χ^2 statistic for large models would generalize to models in our study. Our findings suggest that Gagné and Hancock's approach extends to large models with surprisingly little increase in sample size requirements and that the Swain correction to χ^2 performs fairly well. We argue that likely rejection or model fit should be taken into account when determining sample size requirements and therefore, provide an updated table of minimum sample size that incorporates Gagné and Hancock's method and model fit.

Keywords: confirmatory factor analysis, latent variable reliability, sample size, Swain correction, structural equation modeling

The question of minimum required sample size for use in latent variable modeling is an important issue, particularly when considering the popularity of structural equation modeling (SEM) techniques (Breckler, 1990; Hershberger, 2003; Martens, 2005; Tremblay & Gardner, 1996). Past research recommends a variety of strategies for determining an appropriate sample size (e.g., Anderson & Gerbing, 1984; Boomsma, 1982, 1985; Fan, Thompson, & Wang, 1999; Gerbing & Anderson, 1985; Jackson, 2003; MacCallum, Browne & Sugawara, 1996; MacCallum, Widaman, Zhang, & Hong, 1999; Marsh, Balla, & McDonald, 1988; Tanaka, 1987). Rules of thumb provide some guidance, but can also produce inconsistent estimates (Gagné & Hancock, 2006; Jackson, 2007) that impede the progress of scientific investigation;

Correspondence regarding this article should be addressed to Dennis L. Jackson, Department of Psychology, University of Windsor, 401 Sunset Ave., Windsor, ON, Canada, N9B 3P4. E-mail: djackson@uwindsor.ca

underestimating sample size can lead to incorrect conclusions about the tenability of a model, whereas overestimating can waste researchers' time and resources.

BRIEF OVERVIEW OF SAMPLE SIZE RESEARCH

Rules of thumb can be unconditional, such as establishing a minimum sample size for SEM (e.g., $N = 200$) or conditional, such as basing sample size recommendations on features of the model. An early conditional rule expressed sample size (n) requirements in terms of the number of parameters estimated (q), denoted n/q (see Jackson, 2001, 2003, for more discussion on this rule). This rule of thumb has been endorsed to varying degrees (Bollen, 1989; Herzog & Boomsma, 2009; Kim & Bentler, 2006; Kline, 2005; Marsh et al., 1988; Mueller, 1996; Nevitt & Hancock, 2004; Ullman, 1996); however, recommendations based on n/q have also been questioned (Jackson, 2003, 2007; Marsh, Hau, Balla, & Grayson, 1998).

Such rules of thumb have the limitation in that they do not directly address an important mechanism that influences sample size requirements: the assumption that unique factors are uncorrelated with each other and with common factors (Gorsuch, 1983; MacCallum et al., 1999; MacCallum, Widaman, Preacher, & Hong, 2001). MacCallum and his colleagues (MacCallum et al., 1999; MacCallum et al., 2001) demonstrated that violations of this assumption in samples are attenuated with greater sample size and higher factor loadings (thus higher communalities). Furthermore, the authors spoke to the effect of p/f ratios (ratio of the number of measured variables per latent variable), suggesting that the positive effect of reducing f relative to p is the result of estimating fewer parameters.

Linking sample size requirements to an index of anticipated latent variable reliability is a conditional rule that does address the previously referenced statistical assumption (Gagné & Hancock, 2006). The relationship of sample size requirements to loading size and p/f ratio suggests latent variable reliability as a determinant of necessary sample size. In earlier work, Marsh et al. (1998) manipulated reliability by varying p/f and measured reliability using ω, attributable to McDonald (1985). More recently, Gagné and Hancock (2006) argued that higher communality and p/f should be incorporated into one index in the study of sample size and proper confirmatory factor analysis (CFA) and SEM solutions. Specifically, Gagné and Hancock varied p/f and loading size, expressing their results in terms of ω, as well as another index, H (Hancock & Mueller, 2001), which is a maximal reliability estimate (see Gagné & Hancock for formulas for ω and H and a comparative analysis). Gagné and Hancock found that adequate sample size relates to both ω and H and that model estimates are more trustworthy in conditions where reliability is reasonable and sample size is adequate. They provided a table showing the minimum sample size necessary to maximize the chances of convergence, given H or ω and p/f.

BACKGROUND ON MODEL SIZE RESEARCH

Because all models in Gagné and Hancock's (2006) study had three latent variables, an important question is whether their approach holds for models with a larger number of latent variables. Achieving levels of acceptable fit for larger models might be difficult (Herzog, Boomsma, & Reinecke, 2007; Marsh, Hau, & Grayson, 2005); however, research suggests that increasing the *p/f* ratio can help reduce improper solutions and convergence problems (Anderson & Gerbing, 1984; Boomsma, 1982), but also produces positively biased χ^2 values (Gagné & Hancock, 2006; Kenny & McCoach, 2003; Marsh et al., 1998). Increasing the number of latent and measured variables, but maintaining a constant *p/f* ratio, also increases χ^2 bias (Herzog et al., 2007).

Recently, Herzog et al. (2007) examined the effect of model size with CFA models ranging from 4 to 16 latent variables with three indicators each and found that the maximum likelihood χ^2 statistic is positively biased with large models. However, these authors found that a correction attributed to Swain (1975) showed promise for rescaling the χ^2 value. They argued that studies involving large models are likely rejected due to poor fit and, furthermore, because many commonly used indexes of fit are reexpressions of χ^2 values; this problem is a concern for other fit measures (Herzog & Boomsma, 2009).

THIS STUDY

This current study extends the work of Gagné and Hancock (2006) on sample size and construct reliability to a broader range of CFA models by varying both *p/f* and the number of latent variables (f). Specifically, our extension involves testing a broader range of f than Gagné and Hancock, whose population models all included three latent variables. Increasing the *p/f* ratio and the number of latent variables is expected to produce positively biased χ^2 values; therefore, the second aim of the study extends the work of Herzog et al. (2007) by evaluating the effect of the Swain correction to reduce χ^2 bias across this broader range of models. Because Herzog and Boomsma (2009) found it valuable to correct fit measures using the Swain correction, we also examine corrected and uncorrected versions of fit measures.

METHOD

Design

Consistent with Gagné and Hancock (2006) and Marsh et al. (1998), we used *p/f* levels of 2, 3, 4, 5, 6, 7, and 12; sample sizes of 25, 50, 100, 200, 400, and 1,000; and all factors had population correlation values of .30. There were four levels of f: 3, 6, 12, and 16. Whereas three latent variables were used in previous work, the other levels represent a model that might be frequently encountered in the literature ($f = 6$), and two larger models that would be encountered infrequently ($f = 12$ and 16). Two homogeneous loading conditions were chosen: .40 and .80. This resulted in a 7 (*p/f*) × 4 (f) × 6 (sample size) × 2 (loading size) design, yielding 336 cells. For each cell, we attempted to obtain 1,000 properly converged solutions, but set a maximum of 5,000 replications, as we anticipated that some cells would have high convergence failure rates.

There are other conditions that we could have manipulated. For instance, Gagné and Hancock (2006) found convergence effects for heterogeneous loading conditions. However, to keep a manageable design, we chose to examine only the conditions deemed necessary to adequately answer the central research questions. Notably, our choices resulted in low sample sizes for some model conditions. We anticipated this being problematic and it could be argued that some of the conditions should have been eliminated (e.g., Skrondal, 2000). However, we included these conditions to comment on them (see, e.g., Marsh et al., 1998).

Data Generation and Model Fitting

SAS version 9.1 for Windows was used to generate multivariate normal data according to the population models outlined. All authors created the data generation and fitting programs, based on a template provided by the first author. Every program was checked by one of the other authors. Latent variable variances were set to 1.0, and data were generated from a standard normal distribution, using randomly chosen seed values.

SAS (PROC CALIS) was used to fit the models, using the maximum likelihood estimation. Initial start values for parameter estimates were set to the known population values when fitting models (Gagné & Hancock, 2006). A maximum of 500 iterations and function calls were allowed and if convergence criterion was not satisfied, the models were coded (by SAS) as not converged. For models where the number of variables exceeded the sample size, the ridge option was specified and SAS's default ridge factor was utilized (choosing a ridge factor so that the smallest eigenvalue is approximately 10^{-3}). Parameter estimates, convergence information, and fit values were saved to output files for subsequent analyses.

Data Analysis

Practical significance. We adopted a cutoff of $\omega^2 = .03$ (Anderson & Gerbing, 1984) for interpreting effects. Simple descriptive procedures were used to communicate results such as rejection rates and solution propriety.

Solution propriety. We adopted Gagné and Hancock's (2006) definition of solution propriety as a solution that converged without any improper values (e.g., Heywood cases). We used the phrase proper solution to mean the same thing. Furthermore, we adopted their index of *C,* defined as the number of replications required to obtain 1,000 proper solutions. Satisfactory *C* represents conditions where 1,100 or fewer replications are required to achieve 1,000 proper solutions.

Swain correction. The Swain correction is from an unpublished dissertation (Swain, 1975). Herzog et al. (2007) indicated that asymptotically, the Swain corrected ML χ^2 statistic matches the ML χ^2 distribution. The Swain correction can be defined as:

$$s = 1 - \frac{p(2p^2 + 3p - 1) - q(2q^2 + 3q - 1)}{12dn} \tag{1}$$

where

$$q = \frac{\sqrt{1 + 4p(p+1) - 8d} - 1}{2} \tag{2}$$

For these equations, p is the number of observed variables, d is the degrees of freedom, and n is the sample size, $N - 1$. The adjustment to χ^2 is made by multiplying the ML χ^2 value by the Swain correction s: $\text{ML}_\text{S} = s(\chi^2)$.

Reporting findings. It was not possible to report all of the findings, but descriptive statistics for each condition are available on request. We primarily concentrated on findings that are relevant to our two research questions.

Outcome Expectations

Solution propriety, a function of nonconvergence and improper solutions, could result from empirical underidentification (see, e.g., Rindskopf, 1984; Wothke, 1993); therefore, we expected conditions with low *n/q* ratios to experience high convergence failures. Past research suggested that solution propriety is a function of both *p/f* and loading size, with greater numbers of improper solutions occurring in models with low *p/f* values (such as $p/f = 2$) and low loading conditions (Loading = 0.40). These two conditions should combine to produce high levels of improper solutions. We expected the combination of small sample size, low *p/f,* and small loadings to produce disastrous results.

Given its past performance (Fouladi, 2000; Herzog & Boomsma, 2009; Herzog et al., 2007), we anticipated that the Swain (1975) correction would adequately correct χ^2 values across a broader range of models than previously shown. In addition to biased χ^2 values, we assumed that when models are large relative to sample size, other fit indexes will reflect poorer fit relative to smaller models. Past findings revealed lower Comparative Fit Index (CFI; Bentler, 1990) values and positively biased root mean square error of approximation (RMSEA; Steiger & Lind, 1980) values in smaller sample size conditions (Curran, Bollen, Paxton, Kirby, & Chen, 2002; Ding, Velicer, & Harlow, 1995; Herzog & Boomsma, 2009). Therefore, we expected that applying the Swain (1975) correction to fit indexes would result in more favorable values when models were large, or sample sizes were small as observed by Herzog and Boomsma in the models they studied.

RESULTS

Solution Propriety

Convergence failures occurred more with small sample sizes ($N \leq 100$), larger number of latent variables ($f = 12$ or 16), low loadings (.40), and low *p/f* values ($p/f = 2$ or 3). Table 1 presents an extension of Gagné and Hancock's (2006) table of satisfactory C. Sample sizes in the range of two to four times greater are required for the low loading condition compared to the high loading condition. However, this ratio depends on the level of *p/f.* That is, with higher *p/f* values, the difference in sample size requirement was smaller between the two loading conditions compared to the difference with lower *p/f* values.

Examining χ^2 Bias

As some cells had no or very few proper convergences across the replications, we eliminated the smallest sample size conditions ($N = 25$ and $N = 50$). Taking the first 1,000 observations (proper solutions), we calculated the relative χ^2 bias as suggested by Bandalos (2006). The majority of cells had more than 800 observations, however at the $N = 100$ level there were

TABLE 1
Minimum Sample Size for Satisfactory Convergence ($C \leq$ 1,100)

		N			
p/f	*a*	*3 Factors*	*6 Factors*	*12 Factors*	*16 Factors*
2	.40	—	1,000	1,000	1,000
	.80	400	400	400	400
3	.40	400	400	400	400
	.80	50	50	100	100
4	.40	200	200	200	400
	.80	25	50	50	50
5	.40	200	200	200	200
	.80	25	25	50	50
6	.40	100	100	200	200
	.80	25	25	50	50
7	.40	100	100	100	200
	.80	25	25	25	50
12	.40	50	50	100	100
	.80	25	25	25	50

Note. *p/f* = number of measured variables loading on each factor; *a* = population loading value for all measured variables; *n* = minimum sample size associated with satisfactory convergence; — = cells where $N > 1{,}000$ would be required.

two cells with few observations (12 and 16 factor models, with *p/f* = 2 and population loadings of .40).

As expected, small samples yielded more positively biased ML χ^2 values ($\omega^2 = .065$). Relative χ^2 bias increased with increasing *p/f* ($\omega^2 = .116$) and *f* values ($\omega^2 = .057$). The relative ML_S χ^2 bias was smaller across the cells compared to the uncorrected ML χ^2 values (Table 2). Although it appears the correction helped, there were still some cells that departed considerably from zero, namely the cells with larger *p/f* and *f* values.

There was a significant three-way interaction between *f, p/f,* and sample size ($\omega^2 = .080$). The interpretation of this interaction is straightforward: The degree of χ^2 bias tends to become greater with higher levels of *p/f,* and the rate of increase becomes more pronounced with higher levels of *f;* however, this effect is attenuated as sample size increases. In other words, to reduce χ^2 bias as models grow larger, greater sample sizes are required, even when the Swain (1975) correction is applied.

Rejection Rates

To further explore the ML and ML_S χ^2 bias, we examined the tail behavior of the χ^2 values across the conditions presented in Table 2. Table 3 contains the rejection rates corresponding to the design factors represented in Table 2. As expected, the ML rejection rates for χ^2 increase with increasing levels of *f* and *p/f.* The interaction for *p/f* and *f* can be seen in the tail behavior of the ML χ^2, again with the differences in rejection rates between small and large *p/f* levels being larger with more latent variables (*f*). Thus, correctly specified models are more likely to be rejected at a given sample size as models become larger. The ML_S χ^2 improved rejection

TABLE 2
Relative Bias for Maximum Likelihood and Swain Corrected Maximum Likelihood χ^2 Values by *p/f* and *f*

	ML χ^2 *Bias* *f*					ML_S χ^2 *Bias* *f*				
p/f	*3*	*6*	*12*	*16*	*Total*	*3*	*6*	*12*	*16*	*Total*
2	−0.130	−0.039	0.000	0.016	−0.048	−0.140	−0.060	−0.029	−0.019	−0.070
3	−0.005	0.021	0.060	0.084	0.038	−0.024	−0.014	−0.004	0.000	−0.011
4	0.019	0.042	0.094	0.135	0.071	−0.005	−0.004	0.004	0.010	0.001
5	0.025	0.058	0.136	0.210	0.107	−0.004	0.001	0.012	0.029	0.009
6	0.036	0.073	0.176	0.318	0.151	0.001	0.004	0.022	0.071	0.024
7	0.043	0.087	0.229	0.434	0.198	0.002	0.005	0.036	0.107	0.038
12	0.073	0.174	0.630	0.802	0.420	0.005	0.021	0.131	0.074	0.058
Total	0.010	0.061	0.201	0.306	0.141	−0.023	−0.006	0.028	0.043	0.010

Note. f = the number of latent variables; *p/f* = the number of measured variables per latent variable. The total values represent the mean for each marginal condition. ML = maximum likelihood χ^2; ML_S = Swain corrected maximum likelihood χ^2. The total means and grand mean are not equally weighted across conditions.

rates, but did not completely correct the Type 1 error rates, as sample size is not taken into account. The ML_S χ^2 appears to perform well for most levels of *p/f* for $f = 3$, and for $f = 6$, but deteriorates with larger models. Generally speaking, over more varied conditions, the ML_S χ^2 outperformed the uncorrected ML χ^2, but did not completely correct for model size. Based on findings from the three-way interaction, it is clear that sample size must be taken into account when using the ML_S χ^2 and ML χ^2.

Analysis of Fit Indexes

The RMSEA and the CFI were analyzed in cells with sample sizes of 100 or greater. An exhaustive report of analyses is not undertaken, as the effects are largely what would be

TABLE 3
Rejection Rates for Maximum Likelihood and Swain Corrected Maximum Likelihood χ^2 Values by *p/f* and *f*

	ML Rejection Rates *f*					ML_S *Rejection Rates* *f*				
p/f	*3*	*6*	*12*	*16*	*Total*	*3*	*6*	*12*	*16*	*Total*
2	0.027	0.033	0.051	0.081	0.044	0.027	0.027	0.026	0.027	0.027
3	0.048	0.072	0.272	0.440	0.201	0.041	0.037	0.049	0.057	0.046
4	0.065	0.135	0.495	0.669	0.334	0.050	0.046	0.068	0.124	0.071
5	0.077	0.234	0.660	0.796	0.442	0.049	0.051	0.125	0.303	0.132
6	0.098	0.359	0.754	0.877	0.522	0.053	0.062	0.250	0.354	0.179
7	0.125	0.449	0.826	0.931	0.583	0.054	0.064	0.320	0.424	0.215
12	0.359	0.751	0.995	1.000	0.776	0.059	0.248	0.560	0.556	0.356
Total	0.115	0.295	0.610	0.725	0.429	0.048	0.077	0.211	0.282	0.152

Note. f = the number of latent variables; *p/f* = the number of measured variables per latent variable. The total values represent the rejection rates for each marginal condition. ML = maximum likelihood; ML_S = Swain corrected maximum likelihood.

expected based on either past research or the behavior of the χ^2 statistic just discussed. Our main finding supports Herzog and Boomsma (2009) in that the variability of the Swain corrected versions of RMSEA and CFI were considerably lower than the uncorrected versions of these fit measures across experimental conditions.

Alternative Minimum Sample Size Requirements

Based on the interpretation of the three-way interaction among sample size, *p/f,* and *f,* we constructed an alternative table to the one presented by Gagné and Hancock (2006), and analogous to Table 1 in this article. Table 4 contains minimum sample size requirements for levels of *p/f* and *f,* taking into account solution propriety and expected χ^2 bias. Sample size requirements are presented for each of the combinations of *f, p/f,* and loading size that would be expected to yield either ML or ML_S χ^2 values that fall within the 99% confidence interval of $\alpha = .05$, where the confidence interval was constructed based on Nevitt and Hancock (2004): CI is $[.05 \pm 2.575(.05 \times .95/1000)^{1/2}] \times 100 = [3.23\%, 6.77\%]$. To construct Table 4 we began with the minimum sample size requirement values in Table 1 and determined whether that sample size requirement also yielded an acceptable rejection rate for either the ML χ^2 or ML_S χ^2. If the rejection rate was not acceptable, the next largest sample size was consulted, and so on, until both solution propriety and acceptable rejection rate were satisfied. In some cases the usual ML χ^2 values were acceptable, but with larger model sizes, generally only the ML_S χ^2 values were acceptable.

TABLE 4
Minimum Sample Size for Satisfactory Convergence ($C \leq 1{,}100$) and χ^2 Bias

		N			
p/f	*a*	*3 Factors*	*6 Factors*	*12 Factors*	*16 Factors*
2	.40	—	1,000	**1,000**	**1,000**
	.80	400	400	400	*400*
3	.40	400	400	*400*	*1,000*
	.80	*50*	*50*	*200*	*200*
4	.40	200	*200*	*200*	*400*
	.80	*50*	*100*	*200*	*400*
5	.40	200	*200*	*200*	*400*
	.80	*50*	*100*	*400*	*400*
6	.40	*100*	*100*	*400*	*1,000*
	.80	*200*	*200*	*1,000*	*1,000*
7	.40	*100*	*200*	*1,000*	*1,000*
	.80	*100*	*200*	*400*	*1,000*
12	.40	*200*	*400*	*1,000*	>1,000
	.80	*100*	*1,000*	*1,000*	>1,000

Note. *p/f* = the number of measured variables loading on each factor; *a* = population loading value for all measured variables; *N* = minimum sample size associated with satisfactory convergence. Normal typeface indicates that both ML and ML_S χ^2 values fell within the 99% confidence interval; bold indicates only ML χ^2 fell within the confidence interval; and italics indicates only the ML_S χ^2 fell within the confidence interval.

Finally, Table 1 presents a pattern where sample size requirements are consistently lower for higher loading conditions compared to lower loading conditions. Although this was mostly the case in Table 4, the reverse was true for some larger models. For instance, all of the $p/f = 6$ models in the high loading condition required larger sample sizes than the low loading condition. In these cases, the ML_S χ^2 values for lower sample sizes are close to falling within the 99% confidence interval.

DISCUSSION

This study extends Gagné and Hancock's (2006) work on minimum sample size recommendations based on latent variable reliability and model solution propriety to large models. Our findings demonstrate that this approach to establishing minimum sample size estimates holds promise. In fact, only modest increases in sample size requirements with increasing numbers of latent variables are required (see Table 1). However, we feel that the approach by Gagné and Hancock has an important shortcoming: it does not take into account model fit.

We have incorporated model fit into Gagné and Hancock's (2006) approach by taking into account the tail behavior of the ML and ML_S χ^2 (Table 4). We argue that it is important to consider model fit in the study of minimum sample size requirements because applied researchers do not know the form of the population model for their data (see, e.g., Browne & Cudeck, 1993; MacCallum, 2003). Specifically, while the Gagné and Hancock approach provides minimum sample size requirements resulting in a high likelihood of a proper solution and reasonable parameter estimates, applied researchers might not recognize that they have found a viable model based on the χ^2 statistic and other measures of fit. Take the example of a model with six latent variables and six measured variables per latent variable, and loading values of .80. In consulting Table 1 we assume that a sample size of 25 would be sufficient; however, this model would almost certainly be rejected based on the ML and ML_S χ^2 values (both had rejection rates of 100%), and after an evaluation of the RMSEA (mean value for this cell was .285) and CFI (mean value for this cell was .326). Even the $RMSEA_S$ (mean value for this cell was .126) or the CFI_S (mean value for this cell was .869) causes us to doubt this model. However, a sample size of 200, which is the recommendation in Table 4, would produce a different judgment about this same model. The ML χ^2 and ML_S χ^2 rejection rates were better (.346 and .061, respectively) and both the RMSEA indexes (RMSEA = .018; $RMSEA_S$ = .007) and CFI indexes (CFI = .990; CFI_S = .997) fall within a more acceptable range. The sample size recommendations taking into account model fit (Table 4) result in the high likelihood of a proper solution and the researcher being more likely to recognize the model as having satisfactory fit and, therefore, provide a useful starting point for estimating the required sample size.

The Swain (1975) correction appears to be useful not just when models are made larger by increasing the number of factors, but also when models are made larger by increasing *p/f*. However, as *p/f* and *f* increase, even larger sample sizes are needed, even when applying the ML_S χ^2 statistic, to not overreject true models. Consistent with previous research, many ancillary fit measures can be recalculated using the ML_S χ^2 so that they are not as biased toward model rejection (Herzog & Boomsma, 2009). However, some issues remain unresolved with respect to the Swain (1975) correction. For instance, the correction should be evaluated for different forms of models prior to routine application (Swain).

Limitations and Directions for Future Study

As with other Monte Carlo studies, our investigation involves simplifying decisions that result in lower external validity, such as homogeneous loadings, all latent variables within a model are of the same size (e.g., all $p/f = 3$, or all $p/f = 6$), coarse sample size choices, and the data generated for this study were continuous with a multivariate normal population distribution. Therefore, although encouraging, the promise of this approach should be examined across more varied conditions. Furthermore, we believe that what is important in this study is not as much the table of minimum sample size recommendations, as it is produced under ideal conditions where the true population model is known, but rather the further validation and refinement of a method introduced by Gagné and Hancock (2006) to recommend minimum sample size requirements. We acknowledge that much work still needs to be done. Future research should focus on the shortcomings just outlined, as well as examining the effects of nonnormality.

ACKNOWLEDGMENTS

Preliminary findings from this study were presented at the American Psychological Association annual convention in 2009. Dennis L. Jackson would like to thank Anne Boomsma for comments and suggestions related to this study. The first version of this article was completed while Dennis L. Jackson was on sabbatical visiting the Psychometrics and Statistical Techniques Department at the University of Groningen; we thus wish to acknowledge the collegial welcome and support of members of this department.

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.

Bandalos, D. L. (2006). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course.* Greenwich, CT: Information Age.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

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

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Joreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 149–285). Amsterdam, Netherlands: North-Holland.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 50,* 229–242.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin, 107,* 260–273.

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

Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research, 37,* 1–36.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling, 2,* 119–144.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6,* 56–83.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate normality. *Structural Equation Modeling, 7,* 356–410.

Gagné, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor analysis. *Multivariate Behavioral Research, 41,* 65–83.

Gerbing, D. W., & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research, 20,* 255–271.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future. A festschrift in honor of Karl Jöreskog* (pp. 195–216). Lincolnwood, IL: Scientific Software International.

Hershberger, S. L. (2003). The growth of structural equation modeling: 1994–2001. *Structural Equation Modeling, 10,* 35–46.

Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling, 16,* 1–27.

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling, 14,* 361–390.

Jackson, D. L. (2001). Sample size and the number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling, 8,* 205–223.

Jackson, D. L. (2003). Revisiting sample size and the number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10,* 128–141.

Jackson, D. L. (2007). The effect of the number of observations per parameter in misspecified confirmatory factor analytic models. *Structural Equation Modeling, 14,* 48–76.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10,* 333–351.

Kim, K. H., & Bentler, P. M. (2006). Data modeling: Structural equation modeling. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 161–175). Mahwah, NJ: Erlbaum.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38,* 113–139.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

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

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

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald.* (pp. 275–340). Mahwah, NJ: Erlbaum.

Martens, M. P. (2005). The use of structural equation modeling in counseling psychology research. *The Counseling Psychologist, 33,* 269–298.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum.

Mueller, R. O. (1996). *Basic principles of SEM: An introduction to LISREL and EQS.* New York, NY: Springer.

Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research, 39,* 439–478.

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods and Research, 13,* 109–119.

Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research, 35,* 137–167.

Steiger, J. H., & Lind, J. M. (1980, June). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Swain, A. J. (1975). *Analysis of parametric structures for variance matrices.* Unpublished doctoral dissertation, University of Adelaide, Australia.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58,* 134–146.

Tremblay, P. F., & Gardner, R. C. (1996). On the growth of structural equation modeling in psychological journals. *Structural Equation Modeling, 3,* 93–104.

Ullman, J. B. (1996). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (pp. 709–811). New York, NY: HarperCollins.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. Bollen & S. Long (Eds.), *Testing structural equation models* (pp. 256–293). Newbury Park, CA: Sage.