2017

Preliminary Proactive Sample Size Determination for Confirmatory Factor Analysis Models

Jennifer Koran
Southern Illinois University Carbondale, jennifer.koran@gmail.com

Recommended Citation

Koran, Jennifer. "Preliminary Proactive Sample Size Determination for Confirmatory Factor Analysis Models." *Measurement and Evaluation in Counseling and Development* (Jan 2017). doi:10.1177/0748175616664012.

Preliminary Proactive Sample Size Determination for Confirmatory Factor Analysis Models

Jennifer Koran

Southern Illinois University Carbondale

Author Note

Jennifer Koran, Section on Statistics and Measurement, Southern Illinois University.

The author wishes to thank Dennis L. Jackson for providing data for reanalysis.

Correspondence concerning this article should be addressed to Jennifer Koran, Section on Statistics and Measurement, Department EPSE, 223 Wham Bldg. – MC 4618, Southern Illinois University Carbondale, Carbondale, IL USA 62901-4618. E-mail: jkoran@siu.edu.

Abstract

Proactive preliminary minimum sample size determination can be useful for the early planning stages of a latent variable modeling study to set a realistic scope, long before the model and population are finalized. This study examined existing methods and proposed a new method for proactive preliminary minimum sample size determination.

Keywords: confirmatory factor analysis, sample size, a priori

Preliminary Proactive Sample Size Determination for Confirmatory Factor Analysis Models

The latent variable modeling literature is replete with methodologies for addressing sample size. For example, there are "reactive" methods for statistical inference when data have already been collected and assumptions associated with large sample theory are violated (Marcoulides & Chin, 2013). There are also "proactive" methods for determining minimum sample size prior to collecting data (Marcoulides & Chin, 2013). These latter methods can be further broken down into two types. *Immediate* proactive approaches (often called power analyses) determine minimum sample size necessary to achieve desired statistical goals (such as power) with specified models and hypothesized parameter values. *Preliminary* proactive approaches (often called heuristics) give "ball-park" estimates of minimum sample size necessary to achieve desired statistical goals (typically estimation convergence and proper solutions). It is the latter tack, the preliminary proactive approaches, that researchers often call upon in the early stages of planning a study with the intent of using latent variable modeling. A researcher at this stage may have a particular population and an extensive theory in mind and needs a rough sense of how large a model can realistically be tested in a single study given typical response rates in an accessible population of known finite size. With a general sense of sample size requirements, this researcher could choose to limit the scope of the theory tested in the model or expand the scope of the population to design a study that is both statistically adequate and feasible.

A number of heuristics have been suggested to aid researchers in the preliminary proactive approach to sample size determination for latent variable models. However, the development of these heuristics is based on criteria that have to do with the statistical goals of estimation convergence and unbiased parameter estimates and statistics. The sample size requirements to satisfactorily meet these criteria may be different, perhaps in many cases lower, than those needed to achieve power to detect effects. This latter goal is what most researchers have in mind when they set out to design a study with a latent variable model. However, the heuristics continue to be used in early phases of designing latent variable model research because they are quick and easy to implement. Power analyses, more often used in immediate preliminary approaches, require access to statistical software. What is needed is a preliminary proactive method that determines sample size to achieve multiple statistical goals necessary for successful statistical analysis.

Literature Review

When a researcher in the preliminary phases of designing a study and intending to use a latent variable model asks the question, “What sample size do I need?” desirable outcomes related to three criteria should be considered. First, the estimation converges and produces a proper solution. Second, the model chi-square value (on which many of the most popular model fit statistics are based, i.e., RMSEA) is unbiased, yielding an accurate view of overall model fit. Third, there is sufficient power to detect effects associated with the theory giving rise to the model by evaluating paths between latent constructs. The literature has shown that all three of these considerations are affected by sample size and should be considered when determining a minimum sample size needed for a latent variable modeling study.

It should be noted that in general the empirical studies on sample size in latent variable modeling deal with confirmatory factor analysis (CFA) models in particular. There are two major justifications for this. First, fitting a CFA model in which all latent constructs are freely permitted to covary is a standard first step in the two-step process of fitting structural equation models in practice (Anderson & Gerbing, 1988). Thus, in the early stages of planning structural equation modeling research, planning for a model that will be adequate for this first step is a strong foundation. Second, in terms of model fit directionality among latent constructs does not matter; only the presence or absence of an edge between two nodes matters. Thus, the essential difference in fit between the first step CFA model and the second step structural equation model is primarily the direct effects between latent effects that are fixed to zero in the latter model. In terms of estimation and power to detect an effect, this is very similar to the mechanics of testing a correlation between constructs. With these considerations in mind, the nascent sample size research has largely focused on CFA models and the results have subsequently been generalized in introductory latent variable modeling resources.

Convergence and solution propriety

A number of empirical studies have investigated the minimum sample size needed to obtain unbiased and admissible estimates of model parameters and reduce the chance of convergence failure. Although distinctions can be made, for simplicity throughout the remainder of this article, avoidance of these estimation anomalies will collectively be addressed as the criterion of (model) convergence. Certain absolute minimum sample sizes, regardless of model complexity or other considerations, have been suggested (Anderson & Gerbing, 1984; Boomsma, 1982). Other rules of thumb to arise from studies of estimation stability are generally based on characteristics of model complexity. The number of parameters to be estimated (q; Jackson, 2003) and the ratio of indicators to factors (p/f; Westland, 2010, based on the work of Marsh, Hau, Balla, & Grayson, 1998) have been the basis of heuristics for determining minimum sample size in the context of CFA. Research with CFA models has consistently supported that the number of indicators per factor (*p/f*) is tied to the sample size needed to achieve desirable levels of convergence and proper solutions (Anderson & Gerbing, 1984; Boomsma, 1982; Marsh et al., 1998). Westland (2010) combined these prior research results into a polynomial function $n \geq 50 * ((p/f)^2) - 405 * (p/f) + 1100$ that could be used as a heuristic for minimum sample size.

Unbiased model chi-square statistic

As the size of the model is increased (both in terms of number of factors, *f*, and in terms of indicators per factor, *p/f*), convergence issues are ameliorated, but the model chi-square values become increasingly positively biased (Gagné & Hancock, 2006; Herzog, Boomsma, & Reinecke, 2007; Kenny & McCoach, 2003). However, it has been found that this latter effect is attenuated with larger sample sizes (Boomsma, 1982; Herzog et al., 2007; Jackson, Voth, & Frey, 2013). Thus, the degree of bias in the model chi-square statistic is also an important consideration in determining a priori sample size with CFA models. Herzog et al. (2007) supported the effectiveness of the Swain correction in rescaling the model chi-square statistic to minimize the positive bias. Jackson et al. (2013) applied the Swain correction and then evaluated the sample size necessary to achieve a minimal level of remaining bias in the model chi-square statistic. In some cases, applying this criterion resulted in a higher necessary minimum sample size than consideration of model convergence rates alone.

It should be noted that bias in the model chi-square statistic is not the only bias of concern to researchers. Bias in the parameter estimates and their standard errors are also of interest. These latter types of bias are best investigated in immediate proactive approaches to sample size determination once the researcher has a particular model and parameters of interest on which to focus the analysis.

Power to detect paths between latent constructs

As with other statistical modeling techniques, it is recommended that researchers use an immediate proactive technique to address sample size with an assessment of statistical power for the particular latent variable model in their study design (Hancock, 2006; MacCallum, Browne, & Sugawara, 1996; Muthén & Muthén, 2002; Kim, 2005; Satorra & Sarris, 1985). In the early stages of planning latent variable modeling research, a researcher may not want to invest the time it takes to carry out an immediate proactive power analysis, yet having a sense of what kind of sample size is likely to achieve adequate power would be helpful if power was one criterion as part of a preliminary proactive approach.

Three major approaches to immediate proactive methods for power analysis have been proposed in the literature. The first is the Monte Carlo simulation approach (Muthén & Muthén, 2002). This approach simulates many samples of a given size drawn from a hypothesized population with the hypothesized model fit to data from each sample. Probabilities can then be calculated to express how likely it is that various criteria would be met given a particular sample size. Of course, the probability of rejecting the null hypothesis when it is false is statistical power. The second approach is the RMSEA (root mean square error of approximation) method (MacCallum et al., 1996). This approach assesses the power to obtain an overall model fit, as expressed by the RMSEA statistic value, that is consistent with good model fit. Unfortunately, the latent variable model fit literature is currently fraught with multiplicity and contention, precluding agreement on what constitutes reasonable model fit (see for example, Barrett, 2007;

Bollen & Long, 1993). The third approach is the noncentrality parameter-based method (Hancock, 2006; Satorra & Sarris, 1985). This approach assesses power based on the model comparison underlying a given inferential test of interest, whether this involves many parameters simultaneously or just one.

Sample size considerations based on power are a function of desired power, level of significance, measurement quality, and effect size. Hancock (2006) showed the importance of taking construct reliability into account when conducting sample size calculations for achieving statistical power. Traditional criteria for desired power, level of significance, and effect size should be adequate for the early stages of planning latent variable modeling research, but no preliminary proactive method has been suggested for estimating sample size needed to achieve these conditions.

While the literature has shown that all three of these major considerations – convergence, bias, and power – are affected by sample size, existing proactive preliminary sample size heuristics fall short in taking these considerations into account. Simple sample size heuristics based on the number of observed variables (Nunnally, Bernstein, & Berge, 1967) or the number of model parameters to be estimated (Tanaka, 1987) have been largely discredited for CFA models (Gagné & Hancock, 2006; Marsh et al., 1998; see also Jackson, 2001, 2003), but are promoted in some introductory texts and other sources relied upon by researchers relatively new to latent variable modeling research (Bollen, 1989; Kline, 2011; Mueller & Hancock, 2010). Other texts acknowledge the complexities of latent variable model sample size determination, but are not able to provide solid starting point for applied researchers (Hair, et al., 2006). Rather, the magnitude of factor loadings (sometimes summarized in a measure of construct reliability/replicability) and the number of indicators per factor (*p/f*) seem to be important components in assuring estimation convergence and solution propriety (Anderson & Gerbing, 1984; Boomsma, 1982; Gagné & Hancock, 2006; Marsh et al., 1998; Velicer & Fava, 1987). The size of the model, including the number of factors (*f*) and the number of indicators per factor (*p/f*), is an important consideration in the amount of bias in the model chi-square statistic (Herzog et al., 2007; Jackson et al., 2013; Kenny & McCoach, 2003), and the magnitude of the loadings (as represented in construct reliability) plays an important role in statistical power to detect significant paths between latent constructs (Hancock, 2006). Currently, there are no preliminary proactive minimum sample size heuristics that address all three criteria in the design of CFA studies.

Jackson et al. (2013) applied the first two desired criteria (model convergence and unbiased model chi-square statistic) in producing CFA sample size recommendations. The present study adjusted these minimum sample size recommendations by including the power to detect a medium effect size path between two latent constructs as a third criterion and by changing the criterion for the rate of proper solutions from 90 percent to 99 percent. Second, this study summarized the resulting minimum sample size recommendations in a comprehensive heuristic that is a function of number of factors, number of indicators per factor, and factor loadings, in a spirit similar to a minimum sample size heuristic suggested in Westland (2010). *Please note that for lack of a more fitting term, the word "heuristic" is used here in reference to a helpful procedure for finding an approximate solution*. Finally, the comprehensive heuristic was compared with the observed variables (*p*) heuristic, the model parameters (*q*) heuristic, and the Westland (2010) heuristic. It was hypothesized that the comprehensive heuristic would produce estimates of the minimum sample size that more closely mimicked the sample sizes meeting the three criteria than the other heuristics.

Methods

Data Collection

The present study directly builds upon prior research in Jackson et al. (2013). Thus, the same conditions as in the previous study were investigated in this study. CFA models were used with all factors correlated .3 (medium effect size). The following numbers of factors: 3, 6, 12, and 16, and indicators per factor: 2, 3, 4, 5, 6, 7, and 12, were used in the CFA models. Loadings (factor determinacy) in the models were set to either all .4 or all .8. All three variables were fully crossed for a total of 56 conditions. Maximum likelihood estimation was used.

Convergence rates. Owing to the idea that minimum sample size heuristics are most relied upon by researchers relatively new to latent variable modeling research, the 90 percent convergence rate applied in prior studies (e.g., Gagné and Hancock, 2006; Jackson, 2001, 2003; Jacksonet al., 2013) seemed too low to assure a reasonably smooth model fitting experience. While still allowing for the occasional convergence failure, the convergence rate criterion for this study was set at 99 percent. Model convergence data from Jackson et al. (2013) were reanalyzed to produce recommended sample sizes for 99 percent convergence (with proper solutions).

Unbiased chi-square statistic. The minimum sample size recommendations made in Jackson et al. (2013, p. 93, Table 4) met both a 90 percent convergence criterion and an unbiased model chi-square criterion. The criterion for an unbiased model chi-square was that the expected rejection rate (α) was within the 99 percent confidence interval of .05. The sample size needed for satisfactory convergence was increased until either the uncorrected or the Swain corrected model chi-square rejection rate met this criterion. Thus, where the Jackson et al. (2013, p. 93, Table 4) minimum sample size recommendations were greater than those for 99 percent convergence, this was attributed to the unbiased model chi-square criterion. For this reason Jackson et al.'s (2013, p. 93, Table 4) sample size recommendations were used in the present study to assure that this second criterion was also met.

Power analysis. For each of the 56 conditions, an a priori power analysis following the noncentrality parameter-based method described in Hancock (2006) was conducted to determine the minimum sample size needed to detect a single correlation between factors of .30 (medium effect size) with 80 percent power and a $.05$ level of significance. A medium effect size was chosen because Cohen's (1988) medium effect size recommendations most closely correspond to effects found in published studies in the social sciences. The maximum likelihood fit function was used, and the power analyses were based on the assumption that the data were multivariate normal.

The process for carrying out the power calculations was as follows. First, the desired level of significance ($.05$), the desired power (.80), and the degrees of freedom associated with the number of focal parameters being tested (df=1) were used in a noncentral chi-square distribution calculator to find the associated noncentrality parameter. As all of these considerations are fixed for this study, the noncentrality parameter was 7.85. Second, the effect size and the hypothesized value for the focal parameter were specified. In this study, a correlation of .3 (medium effect size) was used, and the hypothesized value was zero. Third, the values for the peripheral parameters in the model were specified. This varied with the variables being manipulated in the multi-factor CFA models as previously described. Fourth, LISREL 8.80 (Jöreskog & Sörbom, 2006) was used to determine the model-implied covariance matrix for the full model (with the specified effect size for the focal parameter) by fixing the parameter values (see Hancock, 2006). Fifth, the model-implied covariance matrix and the hypothesized value of the focal parameter were used as input in LISREL to determine the value of the model fit function for the reduced model by fixing the parameter values. Finally, the needed sample size was calculated as sample size = 1 + [noncentrality parameter/fit function for reduced model]; for this study this calculation was sample size = 1 + [7.85/fit function for reduced model].

Combining data from three criteria. With recommended sample sizes based on each of three criteria, the sample sizes were compared and the largest minimum sample size recommended for a particular condition was recorded.

Data Analysis

Heuristic development. Line graphs were used to examine trends in the recommended sample size for individual variables thought to be important in predicting the recommended minimum sample size. These variables included number of factors (*f*), number of indicators per factor (*p/f*), loading (*a*), number of observed variables (*p*), number of parameters to be estimated (*q*), and coefficient *H* (maximal reliability; Hancock & Mueller, 2001; Raykov, 2004), which is equivalent to Cronbach's alpha in this study due to all loadings being equal. The most distinctive trend emerged for number of indicators per factor (*p/f*), where the line resembled a *U*-shape with an upper asymptote on both sides as the number of indicators per factor was small or large. Depending on the number of factors (*f*), sometimes an asymptote was not visible in the data, but for all combinations of number of factors (*f*) and loading (*a*), the data was generally consistent with this trend.

The following function was developed to describe the *U*-shaped trend:

$$C = \alpha - \frac{\beta}{\frac{\beta}{\alpha - \gamma} + \left(\frac{p}{f} - \delta\right)^{\varepsilon}} \tag{1}$$

where C is the estimate of the minimum recommended sample size to meet the three criteria, α is the upper asymptote parameter (the same upper asymptote for both the left and right sides), β is a scaling parameter controlling the ratio of length to width of the *U*-shape, γ is the minimum parameter (the height of the bottom of the *U*-shape), *p/f* is the number of indicators per factor, δ is the location of the minimum (the value along the X-axis at the center of the bottom of the *U*-shape), and ε is an exponent that controls the "curviness" of the sides of the *U*-shape.

Values of α, β, γ, δ, and ε were altered for each combination of number of factors (*f*) and loading (*a*) to produce a functional trend that appeared to most closely match the empirical trend in the line graphs. Maximum likelihood estimation was then attempted using these values as starting values. However, owing to a small number of data points, sometimes as low as six data points within a condition, the maximum likelihood estimates produced a function that when graphed was clearly a worse fit to the data than the starting values. Thus, the analysis continued using approximate values produced using trial and error and subsequent examination of the fit of the trend in the line graphs.

Once approximate values of α, β, γ, δ, and ε were determined for each of the 14 combinations of number of factors (*f*) and loading (*a*), each of these five parameters was visually examined across the 14 data points for patterns by number of factors (*f*) and loading (*a*). Functions of number of factors (*f*) and loading (*a*) were developed separately for each parameter to produce the best fit to these data points, typically using OLS regression where appropriate with nonlinear transformations as needed. The new parameter estimates produced using these functions were substituted into equation (1) and again the functional trend was compared against the empirical trend in the line graphs. Adjustments were made and the process repeated until the simplicity of the functions for the parameters and the fit of the functional trends to the empirical trends appeared to be optimized.

Heuristic comparison. Three heuristics recommended for determining minimum sample size in the early stages of planning a CFA study were used for comparison. These heuristics were: ten times the number of observed variables (10p), ten times the number of parameters to be estimated (10q), and Westland's (2010) heuristic (*W1*, $[50 * ((p/f)^2) - 405 * (p/f) + 1100]$). While the comprehensive heuristic was optimized using the data in this study, these three heuristics were not. Thus, comparison also included three more heuristics in which the parameters in the three existing heuristics were estimated with regression models using the generalized method of moments (GMM) to optimize the fit to the recommended minimum sample sizes in this study: a ratio of the number of observed variables (Xp), a ratio of the number of parameters to be estimated (Xq), and a new polynomial function of Westland's (2010) heuristic (*W2*, $[a * ((p/f)^2) + b * (p/f) + c)]$). In all, seven heuristics were compared and the standard error of the estimate was calculated for each. The heuristic with the lowest standard error of the estimate demonstrated the best fit between the recommended minimum sample sizes and the estimated minimum sample sizes from the heuristic.

Results

Recommended minimum sample sizes. With recommended sample sizes based on each of three criteria, the sample sizes were compared. Where the largest minimum sample size recommended for a particular condition was different from that previously recommended in Jackson et al. (2013, p. 93, Table 4), that value was recorded in Table 1. For five of the conditions for either the 99 percent model convergence rate criterion or the unbiased model chi-square criterion, the sample size needed was greater than 1000, but how much above 1000 was unknown. As a specific minimum sample size value was not available, these conditions were not formally included in the model fitting for the comprehensive heuristic. However, the general trend that these values tended to occur in models with the least information (i.e., low loadings and few indicators per factor) and in the largest models (large numbers of factors and the largest numbers of indicators per factor) was taken into consideration when fitting models to the remaining data to develop the heuristic.

Functions for the heuristic parameters. It was found that fixing the exponent $\varepsilon = 8$ allowed for values of Equation (1) that fit all conditions reasonably well. Thus, the exponent ε was fixed to a constant. Functions of number of factors (*f*) and loading (*a*) were found for the remaining four parameters as follows.

$$\alpha = \frac{(f - 9.25)^2}{4 * .18} + 990 \tag{2}$$

$$\beta = 10^{\left[-1.06 + 36.717 * \left(\frac{1}{f}\right) + 6.491 * a - 27.202 * \left(\frac{a}{f}\right)\right]} \tag{3}$$

$$\gamma = 398 + 375 * a + 2a * (f - 4.5)^2 \tag{4}$$

$$\delta = 2.783 + 10.507 * \left(\frac{1}{a * f}\right) \tag{5}$$

The upper asymptote parameter α was found to be best fitted by a wide, flat parabolic function of number of factors (*f*). The scaling parameter β was found to be best fitted by an exponential function of the interaction of loading (*a*) and the inverse of the number of factors (*f*). The minimum parameter γ was found to be best fitted by an interaction of loading (*a*) and a quadratic function of number of factors (*f*). The location parameter δ was found to be best fitted by a linear function of the inverse of the product of number of factors (*f*) and loading (*a*).

Heuristic comparisons. The three heuristics suggested by the literature were fitted to the recommended minimum sample size data using GMM regression to produce the following optimized versions of the heuristics: a ratio of the number of observed variables (8.2p, R^2=.145), a ratio of the number of parameters to be estimated (2.7q, R^2=.176), and a new polynomial function of Westland's (2010) heuristic (*W2*, [6.3$* ((p/f)^2) - 77.4 * (p/f) + 665.9$], R^2=.041). The R^2 values demonstrate that the fit of the three existing heuristics is poor, even after optimizing the model coefficients for the recommended minimum sample size data in the present study.

The estimated sample sizes and standard errors of the estimate for seven minimum sample size heuristics are shown in Table 2. The comprehensive heuristic *C* has the lowest standard error of the estimate among the seven heuristics compared. For models with small numbers of indicators per factor (*p/f*), the ten times the number of observed variables heuristic (*10p*) tended to underestimate the recommended minimum sample size based on the three criteria. As the number of factors in the model (*f*) became large, the ten times the number of parameters heuristic (*10q*) tended to overestimate the recommended minimum sample size based on the three criteria. Westland's (2010) heuristic generally underestimated the recommended minimum sample size based on the three criteria. The optimized versions of these three heuristics performed better, but none of them had a standard error of the estimate close to that of the comprehensive heuristic *C*.

Discussion

This article distinguished between *preliminary* proactive approaches to sample size, which are most useful in the early stages of planning latent variable modeling research to produce general sample size estimates, and *immediate* proactive approaches, which are used to produce a more refined sample size estimate for the specified model and study design. Some authors have suggested that knowledge of immediate proactive approaches is inaccessible to applied researchers (Wolf, Harrington, Clark, & Miller, 2013); I have not generally found this to be true. Rather, I see applied researchers struggling with the poor quality of existing preliminary proactive methods to align the general scope of their research with what might be feasible or realistic to implement. Once a researcher has honed in on a realistic model and population to study, there is only a slight learning curve to acquire a specific technical skill to carry out an immediate proactive method to produce a specific, refined assessment of the sample size needed in a particular study. Immediate proactive analyses are very often carried out using the latent variable modeling software that the researcher plans to use to analyze the data, and typically the software developer has provided a white paper or tutorial to instruct users on how to carry out at least one type of immediate proactive analysis (e.g., Muthén, 2002; Muthén & Asparouhov, 2002). The likely barrier to applied researchers is the idea that a researcher would use this type of computer-intensive technique as a preliminary proactive method, perhaps even before the researcher has finalized a decision regarding the latent variable modeling software he or she plans to use for the analysis.

Results Summary

This study developed recommendations for the minimum sample size necessary to meet three criteria: 99 percent model convergence, minimal bias in the model chi-square statistic, and 80 percent power to detect a medium sized effect in a CFA model in an a priori power analysis. Based on these recommendations a comprehensive preliminary proactive heuristic was developed to estimate the sample size needed to meet these three criteria as a function of number of factors, number of indicators per factor, and magnitude of factor loading in the model. Finally, this study compared the performance of the comprehensive heuristic with previously recommended sample size heuristics used as preliminary proactive approaches.

As expected, the comprehensive heuristic produced estimates of the minimum sample size that were closer to the recommended sample sizes meeting the three criteria than the other heuristics. However, the comprehensive heuristic is also more complicated to implement, requiring a number of calculations. These calculations can be performed with a calculator or spreadsheet software and do not require specialized latent variable modeling software.

The results provide a concrete sense of sample size requirements for researchers in the early stages of planning a CFA study. While Jackson et al. (2013) showed that convergence and bias issues can be avoided with quite small sample sizes (e.g., $\mathrm{n} = 50$) in some small models, the present study showed that this trend is counterbalanced by the need for greater sample size in small models in order to achieve adequate power to detect an effect. However, this study also presented some good news for novice researchers with modest resources for collecting data. Indeed, a straightforward three factor model could be interesting, and the results of the present study as shown in Table 1 suggested that such a study might be carried out with less than 200 cases, provided that the measurement model contains the right combination of indicators per factor and strength of loading. Results for the three factor models with three and six indicators per factor and loadings of .80 were consistent with or larger than the recommendations made by Wolf et al. (2013), which investigated sample sizes meeting multiple criteria in small CFA models (one to three factors).

Assumptions and Limitations

This study made some important assumptions that should be kept in mind when applying the comprehensive heuristic. It was assumed that maximum likelihood was the estimation method. It was also assumed that all variables used in the CFA model are continuous, and that those variables are multivariate normally distributed. These may be reasonable assumptions in the early planning stages of many latent variable modeling studies. If, for example, deeper study of the literature on the phenomenon of interest indicates that some of the variables in the model will not be multivariate normal distributed, then it should be anticipated that a greater sample size will be needed, and the effect of this nonnormality may be taken into account when conducting the immediate preliminary sample size determination for the study. However, if the comprehensive heuristic suggests a sample size that already will be difficult to achieve, then the researcher may consider scaling back the size of the model or making other adjustments while still in the early planning stages.

However, some additional limitations should be considered when applying the comprehensive heuristic and constitute areas for further research and refinement of this heuristic. First, the minimum sample sizes used to develop the heuristic need further refinement, especially for the rates of convergence/proper solutions and rejection rates for the model chi-square. Rates for sample sizes between 400 and 1000, and in some cases above 1000, need to be investigated for several of the conditions. It is possible, for example, that a particular condition may require a sample size of 500 for 99 percent convergence, but the recommended minimum sample size was 1000 because what is known presently is that the minimum sample size is above 400 but less than or equal to 1000. Thus, the comprehensive heuristic for this case may have been developed to generate an estimate closer to 1000, producing an overestimate of the true minimum sample size of 500 needed to meet the criteria. This possibility was taken into consideration when developing the heuristic, but as of right now there is not enough information to know the extent to which the ad hoc interpolation in the heuristic development was accurate.

Likewise, minimum sample sizes for more values of the loading need to be investigated to improve the heuristic. Only two values were used in this study, consistent with values used in previous research, but little can be known about the shape of the relationship with only two values. Again, as of right now there is not enough information to know the extent to which the inherent interpolation in the heuristic development was accurate regarding values of the loading. Application of the comprehensive heuristic with loading values other than .4 and .8 should be interpreted with caution.

In addition, further work adapting the heuristic for conditions in which not all loadings are equal will broaden the applicability of the heuristic. At the same time, adapting the heuristic to use Cronbach's alpha or Coefficient H (maximal reliability; Hancock & Mueller, 2001; Raykov, 2004) instead of the loading (factor determinacy) will improve the practical utility of the comprehensive heuristic. Further investigation of recommended minimum sample sizes for the number of indicators per factor between 8 and 11 and for number of factors between 7 and 11 may help to improve knowledge about the shape of the associations, while at the same time producing additional data points. The additional data points will likely facilitate the use of maximum likelihood to produce more refined and better optimized values for the model parameters to refine the comprehensive heuristic.

Conclusion

While preliminary proactive approaches to sample size should never replace immediate proactive approaches for sample size determination in formal research proposals, the ability to quickly call upon a preliminary proactive method for approximate sample size can be tremendously valuable in the early stages of planning research to set parameters on what is realistic, long before the scope of the model and the scope of the population have been finalized. The impact of this study is to aid in setting realistic parameters on study design while taking into account multiple criteria, including estimation stability, bias, and power in CFA.

The comprehensive heuristic developed in this study represents a compromise between existing tools for planning a latent variable modeling study. On one hand, this approach is more complicated than the simple heuristics that have been previously recommended. On the other hand, this approach can still be implemented relatively easily (i.e., using an Excel file, available from the author upon request) on a computer without specialized latent variable modeling software. Further, the comprehensive heuristic takes into consideration broader criteria than any one of the previously recommended alternatives in the preliminary proactive approach. The results suggest that the comprehensive heuristic is an improvement over previous minimum sample size heuristics and that further refinement could produce an excellent heuristic for use in the early planning stages of a latent variable modeling study.

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. doi: 10.1007/BF02294170

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411-423. http://dx.doi.org/10.1037/0033-2909.103.3.411

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815-824. doi:10.1016/j.paid.2006.09.018

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

Boomsma, A. (1982). Robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction (Part I).* Amsterdam: North Holland.

Cohen, J. (1988). *Statistical power analysis for the social sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gagné, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65-83. doi: 10.1207/s15327906mbr4101_5

Hair, J. F., Black, W., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hancock, G. R. (2006). Power analysis in covariance structure models. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course*. Greenwood, CT: Information Age Publishing.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. H. C. du Toit, & D. Sorbom (Eds.) *Structural equation modeling: Past and present. A Festschrift in honor of Karl G. Jöreskog* (pp. 195-261). Chicago: Scientific Software International.

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 361-390. doi: 10.1080/10705510701301602

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: A Multidisciplinary Journal*, *8*, 205–223. doi: 10.1207/S15328007SEM0802_3

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N: q hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 128-141. doi: 10.1207/S15328007SEM1001_6

Jackson, D. L., Voth, J., & Frey, M. P. (2013). A note on sample size and solution propriety for confirmatory factor analytic models. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*, 86-97. doi: 10.1080/10705511.2013.742388

Jöreskog, K. G., & Sörbom, D. (2006). LISREL 8.80 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 333-351. doi: 10.1207/S15328007SEM1003_1

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*, 368-390. doi: 10.1207/s15328007sem1203_2

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

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. http://dx.doi.org/10.1037/1082-989X.1.2.130

Marcoulides, G. A., & Chin, W. (2013). You write, but others read: Common methodological misunderstandings in PLS and related methods (pp. 31-64). In H. Abdi, W. Chin, V. E. Vinzi, G. Russolillo, & L. Trinchera (Eds.), *New perspectives in partial least squares and related methods*. Berlin, Germany: Springer-Verlag.

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*(2), 181-220. doi: 10.1207/s15327906mbr3302_1

Mueller, R. O., & Hancock, G. R. (2010). Structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 371-383) New York: Routledge.

Muthén, B. O. (2002). *Using Mplus Monte Carlo simulations in practice: A note on assessing estimation quality and power in latent variable models* (Mplus Web Notes: No. 1). http://www.statmodel.com/download/webnotes/mc1.pdf

Muthén, B. O., & Asparouhov, T. (2002). *Using Mplus Monte Carlo simulations in practice: A note on non-normal missing data in latent variable models* (Mplus Web Notes: No. 2). http://www.statmodel.com/download/webnotes/mc2.pdf

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 599-620. doi: 10.1207/S15328007SEM0904_8

Nunnally, J. C., Bernstein, I. H., & Berge, J. M. T. (1967). *Psychometric theory* (Vol. 226). New York: McGraw-Hill.

Raykov, T. (2004). Estimation of maximal reliability: A note on a covariance structure modelling approach. *British Journal of Mathematical and Statistical Psychology*, *57*(1), 21-27. doi: 10.1348/000711004849295

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50,* 83-90. doi: 10.1007/BF02294150

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. http://www.jstor.org/stable/1130296

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*(2), 193-209. doi: 10.1207/s15327906mbr2202_4

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, *9*(6), 476-487. doi:10.1016/j.elerap.2010.07.003

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety.

Educational and Psychological Measurement, 73, 913-934. doi: 10.1177/0013164413495237

Table 1.

Adjusted Minimum Sample Sizes for Satisfactory Convergence, Bias, and Power

		N			
p/f	*a*	3 Factors	6 Factors	12 Factors	16 Factors
2	.40	≥1,129	≥ 1,096	≥ 1,058	1042
	.80	1000	--	--	--
3	.40	1000	613	1000	--
	.80	108	100	--	--
4	.40	450	422	400	--
	.80	98	--	--	--
5	.40	350	324	300	--
	.80	93	--	--	--
6	.40	291	266	--	--
	.80	--	--	--	--
7	.40	251	227	--	--
	.80	--	--	--	--
12	.40	--	--	--	--
	.80	--	--	--	--

Note: -- indicates that the value in this cell is identical to the value in Table 4 of Jackson, Voth, and Frey (2013, p. 93).

Table 2.

Minimum Sample Sizes Estimated by Seven Heuristics for CFA Models with Varying Factors, Indicators, and Loadings

f	*a*	*p/f*	*p*	*q*	*n*	*C*	*10p*	*8.2p*	*10q*	*2.7q*	*W1*	*W2*
3	0.4	3	9	21	1000	754	90	74	210	58	200	491
3	0.4	4	12	27	450	563	120	98	270	74	100	458
3	0.4	5	15	33	350	388	150	123	330	91	100	437
3	0.4	6	18	39	291	293	180	147	390	107	200	429
3	0.4	7	21	45	251	260	210	172	450	123	400	434
3	0.4	12	36	75	200	250	360	294	750	205	2900	647
3	0.8	2	6	15	1000	379	60	49	150	41	400	537
3	0.8	3	9	21	108	169	90	74	210	58	200	491
3	0.8	4	12	27	98	110	120	98	270	74	100	458
3	0.8	5	15	33	93	103	150	123	330	91	100	437
3	0.8	6	18	39	200	102	180	147	390	107	200	429
3	0.8	7	21	45	100	102	210	172	450	123	400	434
3	0.8	12	36	75	100	250	360	294	750	205	2900	647
6	0.4	3	18	51	613	999	180	147	510	140	200	491
6	0.4	4	24	63	422	953	240	196	630	173	100	458
6	0.4	5	30	75	324	565	300	245	750	205	100	437
6	0.4	6	36	87	266	255	360	294	870	238	200	429
6	0.4	7	42	99	227	250	420	343	990	271	400	434
6	0.4	12	72	159	400	1002	720	588	1590	435	2900	647
6	0.8	2	12	39	400	665	120	98	390	107	400	537

6	0.8	3	18	51	100	161	180	147	510	140	200	491
6	0.8	4	24	63	100	102	240	196	630	173	100	458
6	0.8	5	30	75	100	102	300	245	750	205	100	437
6	0.8	6	36	87	200	102	360	294	870	238	200	429
6	0.8	7	42	99	200	139	420	343	990	271	400	434
6	0.8	12	72	159	1000	1004	720	588	1590	435	2900	647
12	0.4	3	36	138	1000	986	360	294	1380	377	200	491
12	0.4	4	48	162	400	428	480	392	1620	443	100	458
12	0.4	5	60	186	300	294	600	490	1860	508	100	437
12	0.4	6	72	210	400	350	720	588	2100	574	200	429
12	0.4	7	84	234	1000	977	840	686	2340	639	400	434
12	0.4	12	144	354	1000	1001	1440	1176	3540	967	2900	647
12	0.8	2	24	114	400	540	240	196	1140	312	400	537
12	0.8	3	36	138	200	191	360	294	1380	377	200	491
12	0.8	4	48	162	200	189	480	392	1620	443	100	458
12	0.8	5	60	186	400	192	600	490	1860	508	100	437
12	0.8	6	72	210	1000	592	720	588	2100	574	200	429
12	0.8	7	84	234	400	969	840	686	2340	639	400	434
12	0.8	12	144	354	1000	1001	1440	1176	3540	967	2900	647
16	0.4	2	32	184	1042	1053	320	262	1840	503	400	537
16	0.4	3	48	216	1000	1006	480	392	2160	590	200	491
16	0.4	4	64	248	400	356	640	523	2480	678	100	458
16	0.4	5	80	280	400	355	800	654	2800	765	100	437
16	0.4	6	96	312	1000	999	960	784	3120	852	200	429
16	0.4	7	112	344	1000	1053	1120	915	3440	940	400	434

16	0.8	2	32	184	400	534	320	262	1840	503	400	537
16	0.8	3	48	216	200	310	480	392	2160	590	200	491
16	0.8	4	64	248	400	310	640	523	2480	678	100	458
16	0.8	5	80	280	400	348	800	654	2800	765	100	437
16	0.8	6	96	312	1000	932	960	784	3120	852	200	429
16	0.8	7	112	344	1000	1046	1120	915	3440	940	400	434
Standard Error of the Estimate						200.5	327.1	314.5	1320.9	308.7	877.3	333.3

Note: *f* is the number of factors; *a* is the factor loading; *p/f* is the number of indicators per factor; *p* is the number of observed variables; *q* is the number of parameters to be estimated; *C* is the comprehensive heuristic; *10p* is the ten times the number of observed variables heuristic; *8.2p* is the optimized ratio of the number of observed variables heuristic; *10q* is the ten times the number of parameters to be estimated heuristic; *2.7q* is the optimized ratio of the number of parameters to be estimated heuristic; *W1* is Westland's (2010) heuristic $50 * ((p/f)^2) - 405 * (p/f) + 1100$; *W2* is the optimized version of Westland's heuristic $6.319 * ((p/f)^2) - 77.433 * (p/f) + 665.891$.