2020

Indicators per Factor in Confirmatory Factor Analysis: More Is Not Always Better

Jennifer Koran
Southern Illinois University Carbondale, jkoran@siu.edu

Indicators per Factor in Confirmatory Factor Analysis: More Is Not Always Better

Jennifer Koran

Southern Illinois University Carbondale, Carbondale, Illinois, U.S.A.

jkoran@siu.edu

Abstract

Although some research in confirmatory factor analysis has suggested that more indicators per factor is generally better, studies have also documented that sample size requirements increase as model size increases. The present study used Monte Carlo simulation to investigate the effect of indicators per factor on sample size requirements. Results demonstrated a nonlinear association between the number of indicators per factor and the minimum required sample size while avoiding six important consequences for the analysis, such as bias in the model chi-square statistic. There is an upper limit for the desirable number of indicators per factor, and this upper limit depends on the number of factors and factor determinacy. The results showed clear patterns for the specific consequences that were most likely with too few or too many indicators per factor and inadequate sample size. Implications for further research are discussed.

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

Indicators per Factor in Confirmatory Factor Analysis: More Is Not Always Better

Past studies in confirmatory factor analysis have suggested that more indicators per factor is generally better (Marsh, Hau, Balla, & Grayson, 1998). However, recent studies have shown that minimum sample size requirements increase as number of indicators per factor increases due to the model size effect (Gagné & Hancock, 2006; Herzog, Boomsma, & Reinecke, 2007; Jackson, Voth, & Frey, 2013; Kenny & McCoach, 2003; Moshagen, 2012; Shi, Lee, & Terry, 2018). As the number of indicators per factor increases, an increase in sample size is needed to control bias in the model chi-square statistic (Boomsma, 1982; Hertzog, Boomsma, & Reinecke, 2007; Jackson, Voth, & Frey, 2013). The need to control bias with a reasonable sample size in practice suggests the existence of an upper limit for the desirable number of indicators per factor in confirmatory factor analysis.

The purpose of this article is to investigate the effect of indicators per factor on minimum sample size requirements in confirmatory factor analysis. The goal is to find the upper limit on the desirable number of indicators per factor beyond which returns diminish due to increasing consequences for the analysis. A Monte Carlo simulation study is presented that aimed to show the nonlinear association between the number of indicators per factor and the minimum required sample size necessary to avoid six important consequences for the analysis. It was hypothesized that there is a number of indicators per factor that results in the smallest required minimum sample size.

Background

There are many potential pitfalls in the practice of conducting confirmatory factor analysis with maximum likelihood estimation. Nonconverged solutions and Heywood cases pose immediate barriers to study completion (Boomsma, 1985; Dillon, Kumar, & Mulani, 1987).

Insufficient power in fitting models leads to indeterminate results and has been linked to replicability problems (Maxwell, 2004). Recently more attention has been given to the insidious problem of bias in the model chi-square statistic, parameter estimates, and standard errors (Jackson, Voth, & Frey, 2013; Muthén & Muthén, 2002; Shi, Lee, & Terry, 2018). There has been concern that bias can exceed ten percent in some circumstances (Muthén & Muthén, 2002), contributing to incorrect inferences about the model or individual parameter effects. While other pitfalls may be more readily identified in practice, the presence of bias in any given analysis largely goes unnoticed.

There is a large body of research investigating factors linked to adverse outcomes in confirmatory factor analysis with maximum likelihood estimation. Six adverse outcomes in confirmatory factor analysis have been tied to inadequate sample size. These include nonconvergence and improper solutions (Boomsma, 1985). Bias in the model chi-square statistic (Herzog, Boomsma, & Reinecke, 2007; Jackson, Voth, & Frey, 2013) and in the parameter estimates and their standard errors (Muthén & Muthén, 2002) have also been shown to be related to inadequate sample size. Finally, inadequate power to detect a misfitting model (MacCallum, Brown, & Sugawara, 1996) and inadequate power for tests of individual parameter estimates (Muthén & Muthén, 2002) are attributed to power being a function of sample size. With such a variety of adverse outcomes tied to inadequate sample size, meeting minimum sample size requirements for confirmatory factor analysis is an important consideration in practice.

Minimum sample size requirements for confirmatory factor analysis have been tied to model complexity. Model complexity includes factor determinacy (saturation) and the model size, characterized as the number of observed variables (see Shi, Lee & Terry, 2018), which can be operationalized in research as a function of the number of factors and the number of indicators per factor. As the model size increases, the required minimum sample size to avoid bias also increases (Jackson, Voth, & Frey, 2013).

The number of common factors to include in a model is largely empirically and theory driven. The number of factors explaining a given set of observed variables may have been suggested by prior exploratory research. On the other hand, a researcher may have in mind a theory to be tested in a latent variable path model for which fitting a confirmatory factor model is a first step in assessing the fit of the measurement portion of the model. In either case, little can be done to change the number of factors without fundamentally changing the purpose of the study.

Several considerations must be taken into account when selecting the number of indicators per factor when specifying a confirmatory factor model. A minimum of three indicators per factor will generally avoid model identification problems (Kenny, 1979; Kenny, Kashy & Bolger, 1998). Construct validity must also be considered, including selecting indicators that will maintain content coverage and avoid method bias (Lambie, Blount & Mullen, 2017; Podsakoff, MacKenzie & Podsakoff, 2012). It is important to have an adequate number of indicators to achieve a desirable level of factor reliability or replicability (Hancock & Mueller, 2001; Raykov, 2004). Beyond these basic considerations, researchers may have considerable latitude in decisions about the number of indicators per factor. Such decisions are informed by theories of best practice based on empirical research studies of sample size, solution propriety, bias, and power.

In theories of best practice, it has been suggested that more indicators per factor is generally better in confirmatory factor analysis (Marsh, Hau, Balla, & Grayson, 1998). However, more indicators per factor will not necessarily be better because it leads to increases in the minimum required sample size. If this increase in the minimum required sample size is ignored, there is a greater chance that the obtained sample size will be inadequate and adverse outcomes will occur in confirmatory factor analysis. It would be better to have a number of indicators per factor that minimizes sample size requirements while avoiding adverse outcomes. The purpose of this study is to address four research questions that would reconcile the indicators per factor literature with the sample size literature.

1. Is there an upper bound for the desirable number of indicators per factor based on considerations of sample size, solution propriety, bias, and power?
2. What is the role of solution propriety, bias, and power in determining the minimum required sample size?
3. How does the minimum required sample size differ when there is heterogeneous factor determinacy?
4. What is the effect of adding a weak indicator on the minimum required sample size?

The scope of the study excluded considerations of model identification, measurement reliability, and validity that are well established in the literature and instead focused on further considerations of sample size, solution propriety, bias, and power that need to be reconciled to inform best practice when selecting the number of indicators per factor.

Method

Addressing these questions required a Monte Carlo design that built upon prior knowledge about the connection between indicators per factor and sample size while overcoming some of the limitations of prior Monte Carlo study designs. While mathematical exposition is superior to simulation, some adverse outcomes, such as nonconvergence, have yet to be generally explained by mathematical theory (Marsh, Hau, Balla, & Grayson, 1998).

Population Models

The present study used similar conditions as in prior studies (Gagné & Hancock, 2006; Jackson, Voth, & Frey, 2013; Marsh, Hau, Balla, & Grayson, 1998). Confirmatory factor models with all factors correlated .3 (medium effect size) were used to generate the data (Jackson, Voth, & Frey, 2013). Because the prior literature indicated that the number of factors was an important facet of model complexity, the number of factors in the model was varied with three, six, or twelve factors (Jackson, Voth, & Frey, 2013). Factor determinacy was also varied with standardized loadings set to all .4, all .6, or all .8 (Gagné & Hancock, 2006). A minimum of three indicators per factor was used so that all models were over-identified and not subject to complications due to empirical under-identification. Ten levels of indicators per factor from 3 to 12 were adopted, consistent with levels used in prior studies (Marsh, et al., 1998; Gagné & Hancock, 2006; Jackson, Voth, & Frey, 2013), to provide a clear picture of the shape of the trajectories. Number of factors, factor determinacy, and number of indicators per factor were fully crossed for 90 base conditions used to address Research Questions 1 and 2.

To address Research Question 3 regarding how the minimum required sample sizes differ when there is heterogeneous factor determinacy, 16 conditions were added in which loading magnitudes were varied within each factor but had the same average loading magnitude as in the comparable base condition with three or six factors and .4 or .6 standardized loadings. Four levels of indicators per factor from three to six were used in these added conditions. To address Research Question 4 regarding the effect of adding a weak indicator on the minimum required sample size, 16 conditions were added in which a single indicator of .2 was added to each factor from the comparable base condition with three or six factors and 4 or .6 standardized loadings. Four levels of indicators per factor from four to seven were used in these added conditions.

Criteria

Sample size was considered adequate in this study if six criteria were met. These criteria were established to minimize chances of six different adverse outcomes for confirmatory factor analysis based on the existing sample size literature. First, to minimize the chance of inadequate power to detect overall model fit, the sample size had to yield 80 percent power at a .05 level of significance to reject a null hypothesis of good fit (RMSEA = .05) when presented with a model with not-good fit (RMSEA = .08; Browne & Cudeck, 1993). The analytical power calculation based on the test of close fit (MacCallum, Brown, & Sugawara, 1996) was performed prior to data generation.

Second, to minimize the chance of improper solutions, the sample size had to produce a minimum of 99 percent proper solutions across replications. This differed from prior studies, which used either 90 percent proper solutions (Gagné & Hancock, 2006; Jackson, Voth, & Frey, 2013) or 100 percent proper solutions (Wolf, Harrington, Clark, & Miller, 2013). The 99 percent cut value was chosen to provide a high level of assurance that a given sample size would yield a proper solution while allowing for a trivial proportion of convergence anomalies due to sampling variability across the multiple replications of the simulation study. Trivial numbers of improper solutions were subsequently deleted prior to computations associated with the remaining criteria.

Third, to minimize the chance of excessive bias in the model chi-square statistic, the sample size had to have a reasonable rejection rate for the model chi-square statistic (Jackson, Voth, & Frey, 2013). In addition to the uncorrected model chi-square statistic, the Swain corrected model chi-square statistic was calculated by multiplying the model chi-square statistic by the following correction factor

$$s = 1 - \frac{p(2p^2+3p-1)-q(2q^2+3q-1)}{12(df)(n-1)}, \quad (1)$$

where

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

and p is the number of manifest variables in the model (Hertzog, et al., 2007). For the model chi-square statistic to be considered unbiased, the rejection rate across replications for either the uncorrected or the Swain corrected model chi-square statistic had to be within the 99 percent confidence interval of .05 (99% CI: $.05 \pm 2.575(.05 * .95/1000)$^½ $= [0.0323, 0.0677]$; Nevitt & Hancock, 2004).

Fourth, to minimize the chance of bias in the parameter of interest, the sample size had to produce a maximum of ten percent average relative bias in the parameter estimate (Muthén & Muthén, 2002). For the purpose of this study the correlation between the first two factors, $\hat{\phi}$, was selected as the parameter of interest. The relative bias in the estimate of the correlation between the first two factors, $\hat{\phi}$, was computed as

$$RB_{\hat{\phi}} = \frac{\bar{\hat{\phi}}-.30}{.30}. \tag{3}$$

where $\bar{\hat{\phi}}$ is the average of the estimates of the factor correlation across replications.

Fifth, to minimize the chance of bias in the standard error of the parameter of interest, the sample size had to produce a maximum of five percent average relative bias in the standard error of the factor correlation (Muthén & Muthén, 2002). The relative bias in the standard error of the correlation between the first two factors, $\widehat{SE}$, was computed as

$$RB_{\widehat{SE}} = \frac{\overline{\widehat{SE}}-SD(\hat{\phi})}{SD(\hat{\phi})}, \tag{4}$$

where $\overline{\widehat{SE}}$ is the average of the estimates of the standard error of the factor correlation across replications and $SD(\hat{\phi})$ is the standard deviation of the parameter estimates of the factor correlation, $\hat{\phi}$, across replications (Muthén & Muthén, 2002).

Sixth, to minimize the chance of inadequate power to detect statistical significance of the parameter of interest, the sample size had to have a minimum 80 percent power to detect a .30 (medium effect size; Cohen, 1988) factor correlation with a .05 level of significance (Muthén & Muthén, 2002). In each replication the null hypothesis was rejected if

$$\frac{\hat{\phi}}{\widehat{SE}} > 1.96. \quad (5)$$

Power was calculated as the percent of replications in which the null hypothesis that the factor correlation is zero was rejected. Again, all six criteria had to be met for a sample size to be considered adequate for the purposes of this study. The criteria were selected to attain a high probability of a successful analysis unhindered by the major concerns discussed in the confirmatory factor analysis literature: nonconvergence, solution impropriety, excessive bias, or low power.

Procedure

For each condition multivariate normal data with standard deviations drawn from *N*(5,1) were generated in SAS 9.4 (SAS Institute Inc., 2018) according to the population models. Maximum likelihood estimation in LISREL 8.80 (Jöreskog & Sörbom, 2006) was used to fit the population model to the simulated data, using population values as starting values. The use of population values as starting values contributed to the generalizability of the results to other latent variable modeling software (Gagné & Hancock, 2006). Convergence information, the model chi-square statistic, and standardized parameter estimates and their corresponding standard errors were saved for further analysis in SAS.

To model accurately the association between the number of indicators per factor and the minimum required sample size, it was necessary to obtain reasonably precise estimates of minimum required sample size. The sample sizes tested were tailored to each condition in the study. For example, in the base condition with six factors, .8 standardized loadings, and five indicators per factor, sample sizes of 75, 100, 125, and 150 were tested, whereas in the base condition with six factors, .4 standardized loadings, and three indicators per factor, sample sizes of 625, 700, 750, 850, 875, and 900 were tested. This approach avoided wasting computational resources with sample sizes that were much too large or much too small based on prior research.

In each condition an initial sample size to be tested was selected based on estimates from the results of Jackson, Voth, and Frey (2013), from applying the Satorra and Saris (1985) method for power to detect a .30 factor correlation, and from applying the MacCallum, Brown, and Sugawara (1996) test of not-close fit method for power to detect overall model fit. For a given sample size 10,000 replications (Muthén & Muthén, 2002) were conducted. The results were evaluated with the six criteria for adequate sample size. Depending on whether all six criteria were met and how close the results were to meeting all six criteria, the sample size was adjusted, and 10,000 replications were conducted at the new sample size. When the adequate sample size was found to be within an interval of 25 observations (e.g., if a sample size of 450 did not meet all criteria but 475 did), the larger value on the interval was recorded as the final value for the minimum required sample size for that condition.

Results

A total of 3,040,000 Monte Carlo solutions were analyzed to investigate the association between the number of indicators per factor and the minimum required sample size. The minimum required sample size for the base conditions can be found in Table 1. As the number of indicators per factor increased, most conditions showed a pattern in which sample sizes decreased at first, but then hit a turning point after which sample sizes increased again. As illustrated in

Figure 1, this U-shaped pattern was evident among the levels of indicators per factor studied for six out of nine combinations of number of factors crossed with factor determinacy.

Insert Table 1 and Figure 1 about here

Thus, in addressing Research Question 1, there does appear to be an upper bound for the desirable number of indicators per factor. However, the number of indicators per factor marking the turning point in the U-shaped curve differed based on the number of factors and factor determinacy. For example, with six factors and .4 standardized loadings, nine indicators per factor is desirable because it results in the lowest minimum required sample size as shown in Table 1. However, with six factors and .6 standardized loadings, the desirable number of indicators per factor drops to six. However, if we double the number of factors to 12 factors with .6 standardized loadings, four is the desirable number of indicators per factor.

With regard to Research Question 2, the criteria driving the minimum required sample size at each point, called the driving factors, may be considered. As seen in Figure 1, prior to the turning point in the U-shaped curve, the driving factors tended to be adequate percentage of proper solutions, adequate power for the RMSEA test of close fit (MacCallum, Brown, & Sugawara, 1996), and a reasonable level of average relative bias in the standard error of the parameter estimate (Muthén & Muthén, 2002). Close to the turning point in the curve, adequate power to detect a medium size effect for the parameter of interest (Muthén & Muthén, 2002) tended to also be a driving factor. Beyond the turning point in the U-shape, having a reasonable rejection rate for the model chi-square statistic (Nevitt & Hancock, 2004) was the driving factor for the minimum required sample size. The relative bias of the parameter of interest was not a driving factor in any of the conditions studied.

The three combinations that did not exhibit a U-shaped pattern appear to be extreme cases due to the following main effects. First, as factor determinacy increased, required sample sizes generally decreased. This was an expected effect due to indicators with higher loadings on a factor providing information that is more reliable and thus requiring fewer cases for parameter estimation and statistical power. Exceptions occurred in models with the largest number of observed variables where having a reasonable rejection rate for the model chi-square statistic (Nevitt & Hancock, 2004) was the driving factor for the minimum required sample size. Second, as the number of factors increased, required minimum sample sizes generally increased. This was also expected due to the model size effect. Exceptions occurred with relatively weak loadings and few indicators per factor where having a reasonable level of average relative bias in the standard error of the parameter estimate (Muthén & Muthén, 2002) was the driving factor.

Notice that the three combinations that did not exhibit the U-shape had extreme values on these factors. In the case with three factors and .4 standardized loading (shown in the upper left panel of Figure 1), a small model with relatively weak indicators, extremely large sample sizes were required to have a reasonable level of average relative bias in the standard error of the parameter estimate. Increasing the number of indicators per factor increased estimation stability up to the maximum of 12 indicators studied, consistent with a demonstration of the left hand side of the U-shape. It is likely that somewhere beyond 12 indicators having a reasonable rejection rate for the model chi-square statistic would become the driving factor and an increase in required minimum sample size would be needed, consistent with the pattern observed for other combinations.

In cases with 6 or 12 factors and .8 standardized loading (shown in the lower middle and lower right panels of Figure 1), relatively large models with strong indicators, the opposite pattern was found. Required minimum sample sizes were at a minimum at three indicators per factor, consistent with a demonstration of the right hand side of the U-shape. In the case with six factors and .8 standardized loading, the driving factor for the minimum required sample size was adequate power for the RMSEA test of close fit (MacCallum, Brown, & Sugawara, 1996). In the case with 12 factors and .8 standardized loading, the driving factor for the minimum required sample size was having a reasonable rejection rate for the model chi-square statistic (Nevitt & Hancock, 2004). For both cases with 6 or 12 factors and .8 standardized loading, power and the need to control bias in the model chi-square statistic resulted in increases in the required minimum sample size for any number of indicators per factor beyond three. This result suggests that for large models with strong indicators there is little benefit in having more than three indicators per factor.

The minimum required sample size for each condition addressing Research Question 3 can be found in Table 2. Results for comparable homogeneous loading base conditions are also included on the left hand side of Table 2 for ready comparison. When the standardized loadings were all .5 or above (average loading .6), heterogeneity in the loadings made little difference in the minimum required sample size. However, when the standardized loadings were all .5 or below (average loading .4), results became more sensitive to the exact configuration of loadings. Common to all of these conditions was the driving factor of the relative bias in the standard error of the estimate of the factor correlation. In the larger models (six factors), the heterogeneous loadings reduced sample size, but in the smaller models (three factors) the heterogeneous loadings reduced only the most extreme sample size found in the base conditions. It appears that a small model with weak loadings creates a situation where the relative bias in the standard error of the estimate of the factor correlation will be very sensitive to the exact configuration of loadings. Regardless, the same general U-shaped pattern of sample size by indicators per factor was found in the heterogeneous loading conditions as in the base conditions.

Insert Table 2 about here

With regard to Research Question 4, the minimum required sample size for each base condition and the corresponding condition with an additional weak indicator can be found in Table 3. All base conditions were at a number of indicators per factor lower than the desirable number, except for the six factor all .6 loading condition with six indicators per factor. In most cases the addition of the weak indicator reduced or maintained the minimum required sample size. One of the few exceptions was the condition with three factors, four indicators per factor, and .4 standardized loadings. When a single weak indicator was added, the sample size increased. Consistent with results found with the heterogeneous indicators, the driving factor in this case was relative bias in the standard error of the estimate of the factor correlation. This provides further evidence that a small model with relatively weak loadings creates a situation where the relative bias in the standard error of the estimate of the factor correlation will be very sensitive to the exact configuration of factor loadings. When the base condition was at the desirable number of indicators, as in the six factor all .6 loading condition with six indicators per factor, the minimum required sample size increased as expected due to the model size effect requiring a larger sample size to maintain a reasonable rejection rate for the model chi-square statistic.

Insert Table 3 about here

Discussion

The results of this study showed a nonlinear association between the number of indicators per factor and the minimum required sample size necessary to avoid six important consequences for the analysis. In most of the cases studied, an upper limit was reached for the number of indicators per factor associated with the smallest value for the minimum required sample size. Beyond this point larger minimum sample size was necessary to compensate for the increasing model size brought about by the additional indicators per factor. It was demonstrated that this increase was largely due to bias in the model chi-square statistic. However, other factors, namely adequate power for the parameter of interest and reasonable average relative bias in the standard error of the parameter estimate, also play an important role near the turning point where the maximum desirable number of indicators is reached. While the exact value for the desirable number of indicators per factor depended on model size and factor determinacy, this study demonstrated the existence of a practical upper bound on the number of indicators per factor.

Models with relatively weak loadings tended to have a larger desirable number of indicators per factor than models with stronger loadings based on considerations of sample size, solution propriety, bias, and power. Models with few factors also tended to have a larger desirable number of indicators per factor than models with more factors. Models with many factors and strong indicators reached a minimum required sample size with as few as three indicators per factor.

Limitations

This study had some important limitations that should be considered in interpreting the results. First, some of the six criteria are not perfectly consistent. For example, it is known that the bias in the Swain-corrected model chi-square is only approximately consistent, and not perfectly consistent. Thus, minimum sample sizes meeting this criterion can only be approximately determined, but not perfectly determined.

Continuing developments may alter the minimum sample sizes reported in this study. For example, some evidence suggests that other correction factors, such as Yuan, Tian, and Yanagihara's (2015) empirically corrected chi-square, may be more effective than the Swain correction in reducing bias in the model chi-square statistic (Shi, Lee & Terry, 2018). This may reduce minimum sample sizes and increase the number of indicators per factor beyond which diminishing returns are found. However, improved corrections may not alter the basic implication of this study that more indicators per factor is not necessarily better. Bias in the model chi-square with large numbers of observed variables has been shown to persist across corrected statistics in spite of large sample sizes (Yang, Jiang, & Yuan, 2018).

As a single experiment with limited conditions, this study provided insight into the exact number of desirable indicators per factor for only a limited number of the wide variety of circumstances encountered in practice. It should not be assumed that the results of this study would generalize to indicators with nonnormal distributions, such as ordinal items, or estimation methods where bias in the model chi-square is not a concern, such as Bayesian estimation. It has recently been proposed that overall model power and bias in the chi-square statistic be investigated concurrently while accounting for data nonnormality (Yuan, Zhang & Zhao, 2017). Future research should consider sample size implications for conditions in which the normality assumption is violated. Further research into this phenomenon should yield insights into the desirable number of indicators per factor for a comprehensive range of conditions.

Implications

Taken at face value, the notion that there is a "sweet spot" for the number of indicators per factor may seem to contradict the findings from Marsh, Hau, Balla, and Grayson (1998), from which many have concluded that more indicators per factor is better. While Marsh et al. briefly acknowledged the growing bias in the model chi-square as the size of the model increased, the study did not control for this bias. Jackson, Voth, and Frey (2013) later introduced controlling bias in the model chi-square as a criterion for minimum sample size. The present study has built upon both Marsh, et al. and Jackson, et al. to demonstrate that more indicators per factor is not always better. Increasing the number of indicators per factor beyond the desirable number requires that the researcher either seek a larger sample or accept demonstrated consequences for the analysis, such as bias in the model chi-square statistic. Researchers should be cautioned against feeling a need to include as many indicators as possible, following the more-is-better philosophy that was derived from the prior literature. The more-is-better approach to the number of indicators per factor combined with failure to consider bias in the model chi-square statistic can easily lead to situations in which sample size is inadequate and conclusions erroneous.

The results of this study provide new insight into the question of whether there is a compensatory relation between sample size and number of indicators per factor. That is, can having a large number of indicators per factor help compensate for small sample size and a large sample size help compensate for few indicators per factor, as claimed by Marsh, Hau, Balla, and Grayson (1998)? Based on the nonlinear relation found in this study, the answer is that there can be such a compensatory relation, but only within certain parameters. With a model of three factors and relatively weak loadings, having as many as 12 indicators per factor can reduce the minimum required sample size to a reasonable level of approximately 200, while having an extraordinarily large sample size of over 1,000 will likely provide reasonable outcomes with as few as three indicators per factor. However, the results of this study suggest that for models with large numbers of factors and relatively strong loadings, a large sample size will instead be necessary to compensate for having *many* indicators per factor, not fewer indicators per factor.

An interesting area for future research is the situation where factor determinacy differs across factors in a model. It is common to have models in which some constructs generally have more reliable indicators than other constructs. Related to this problem is the effect of differing numbers of indicators per factor across factors on the minimum required sample size for the entire model. Addressing variation within a model requires a number of considerations that were clearly beyond the practical scope of the current study.

In conclusion, the nonlinear association between the number of indicators per factor and the minimum required sample size found in this study demonstrates that more indicators per factor is not always better. Once the upper limit on the number of indicators per factor is reached, increasing minimum sample size is necessary to compensate for the increasing model size brought about by the additional indicators per factor. Thus, the results contradict the universal applicability of a compensatory relation between sample size and number of indicators per factor.

References

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.

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

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin*, *101*, 126-135.

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.

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 (pp. 195-216). Lincolnwood, IL: 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.

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.

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

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S., Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol.1, 4th ed., pp. 233-265). Boston: McGraw-Hill.

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.

Lambie, G. W., Blount, A. J., & Mullen, P. R. (2017). Establishing content-oriented evidence for psychological assessments. *Measurement and evaluation in counseling and development*, *50*(4), 210-216.

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.

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.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, *9*, 147-163.

Moshagen, M. (2012). The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling, 19*, 86–98. doi:10.1080/10705511.2012.634724

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.

Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research*, *39*(3), 439-478.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology*, *63*, 539-569.

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.

SAS Institute Inc. (2018). *SAS/STAT 15.1 user's guide*. Cary, N.C.: SAS Institute Inc.

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

Shi, D., Lee, T., & Terry, R. A. (2018). Revisiting the model size effect in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 21-40.

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.

Yang, M., Jiang, G., & Yuan, K. H. (2018). The performance of ten modified rescaled statistics as the number of variables increases. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 414-438.

Yuan, K. H., Tian, Y., & Yanagihara, H. (2015). Empirical correction to the likelihood ratio statistic for structural equation modeling with many variables. *Psychometrika, 80*, 379–405. doi:10.1007/s11336-013-9386-5

Yuan, K. H., Zhang, Z., & Zhao, Y. (2017). Reliable and more powerful methods for power analysis in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 315-330.

Table 1

Minimum Required Sample Size by Number of Indicators per Factor with Homogeneous Loadings

		Indicators per factor									
f	Loading	3	4	5	6	7	8	9	10	11	12
3	0.4	1050	500	550	375	300	275	250	275	225	200
3	0.6	375	211	175	150	150	125	125	125	150	150
3	0.8	375	211	145	110	100	100	100	125	150	150
6	0.4	900	625	500	450	275	300	225	250	275	350
6	0.6	225	175	175	150	175	200	225	275	325	350
6	0.8	117	125	125	150	200	200	250	275	325	350
12	0.4	675	475	500	350	350	500	600	600	750	850
12	0.6	250	175	275	325	375	500	625	675	750	850
12	0.8	125	200	300	350	375	525	625	675	750	875

Note. *f* = number of factors. Loading is standardized.

Table 2

Minimum Required Sample Size for Homogeneous Versus Heterogeneous Loadings

			Homogeneous			Heterogeneous		
f	Average loading	Indicators per Factor	Loadings	Minimum n	Driving factor	Loadings	Minimum n	Driving factor
3	0.4	3	.4, .4, .4	1050	RBse	.3, .4, .5	850	RBse
3	0.4	4	.4, .4, .4, .4	500	RBse	.3, .3, .5, .5	625	RBse
3	0.4	5	.4, .4, .4, .4, .4	550	RBse	.3, .3, .4, .5, .5	550	RBse
3	0.4	6	.4, .4, .4, .4, .4, .4	375	RBse	.3, .3, .3, .5, .5, .5	350	RBse
3	0.6	3	.6, .6, .6	375	RMSEA	.5, .6, .7	375	RMSEA
3	0.6	4	.6, .6, .6, .6	211	RMSEA	.5, .5, .7, .7	211	RMSEA
3	0.6	5	.6, .6, .6, .6, .6	175	RBse & power	.5, .5, .6, .7, .7	150	power
3	0.6	6	.6, .6, .6, .6, .6, .6	150	power	.5, .5, .5, .7, .7, .7	150	power
6	0.4	3	.4, .4, .4	900	RBse	.3, .4, .5	675	RBse
6	0.4	4	.4, .4, .4, .4	625	RBse	.3, .3, .5, .5	400	RBse & power
6	0.4	5	.4, .4, .4, .4, .4	500	RBse	.3, .3, .4, .5, .5	400	RBse
6	0.4	6	.4, .4, .4, .4, .4, .4	450	RBse	.3, .3, .3, .5, .5, .5	300	RBse
6	0.6	3	.6, .6, .6	225	RBse & power	.5, .6, .7	200	RBse & power
6	0.6	4	.6, .6, .6, .6	175	RBse & power	.5, .5, .7, .7	175	power
6	0.6	5	.6, .6, .6, .6, .6	175	power	.5, .5, .6, .7, .7	175	RBse
6	0.6	6	.6, .6, .6, .6, .6, .6	150	ChiSq & power	.5, .5, .5, .7, .7, .7	175	ChiSq

Note. *f* = number of factors. Loadings are standardized. RMSEA = adequate power for the RMSEA test of close fit; RBse = reasonable average relative bias in the standard error of the parameter estimate; power = adequate power for the parameter of interest; ChiSq = reasonable rejection rate for the model chi-square statistic.

Table 3

Minimum Required Sample Size with the Addition of a Single Weak Indicator

		Base condition			Add one weak indicator		
f	Indicators per Factor	Loadings	Minimum n	Driving factor	Loadings	Minimum n	Driving factor
3	3	.4, .4, .4	1050	RBse	.4, .4, .4, .2	775	RBse
3	4	.4, .4, .4, .4	500	RBse	.4, .4, .4, .4, .2	725	RBse
3	5	.4, .4, .4, .4, .4	550	RBse	.4, .4, .4, .4, .4, .2	450	RBse
3	6	.4, .4, .4, .4, .4, .4	375	RBse	.4, .4, .4, .4, .4, .4, .2	375	RBse
3	3	.6, .6, .6	375	RMSEA	.6, .6, .6, .2	225	RBse & power
3	4	.6, .6, .6, .6	211	RMSEA	.6, .6, .6, .6, .2	175	RBse & power
3	5	.6, .6, .6, .6, .6	175	RBse & power	.6, .6, .6, .6, .6, .2	150	power
3	6	.6, .6, .6, .6, .6, .6	150	power	.6, .6, .6, .6, .6, .6, .2	150	power
6	3	.4, .4, .4	900	RBse	.4, .4, .4, .2	675	RBse
6	4	.4, .4, .4, .4	625	RBse	.4, .4, .4, .4, .2	600	RBse
6	5	.4, .4, .4, .4, .4	500	RBse	.4, .4, .4, .4, .4, .2	400	RBse
6	6	.4, .4, .4, .4, .4, .4	450	RBse	.4, .4, .4, .4, .4, .4, .2	450	RBse
6	3	.6, .6, .6	225	RBse & power	.6, .6, .6, .2	225	power
6	4	.6, .6, .6, .6	175	RBse & power	.6, .6, .6, .6, .2	175	RBse & power
6	5	.6, .6, .6, .6, .6	175	power	.6, .6, .6, .6, .6, .2	150	ChiSq & power
6	6	.6, .6, .6, .6, .6, .6	150	ChiSq & power	.6, .6, .6, .6, .6, .6, .2	175	ChiSq

Note. *f* = number of factors. Loadings are standardized. RMSEA = adequate power for the RMSEA test of close fit; RBse = reasonable average relative bias in the standard error of the parameter estimate; power = adequate power for the parameter of interest; ChiSq = reasonable rejection rate for the model chi-square statistic.

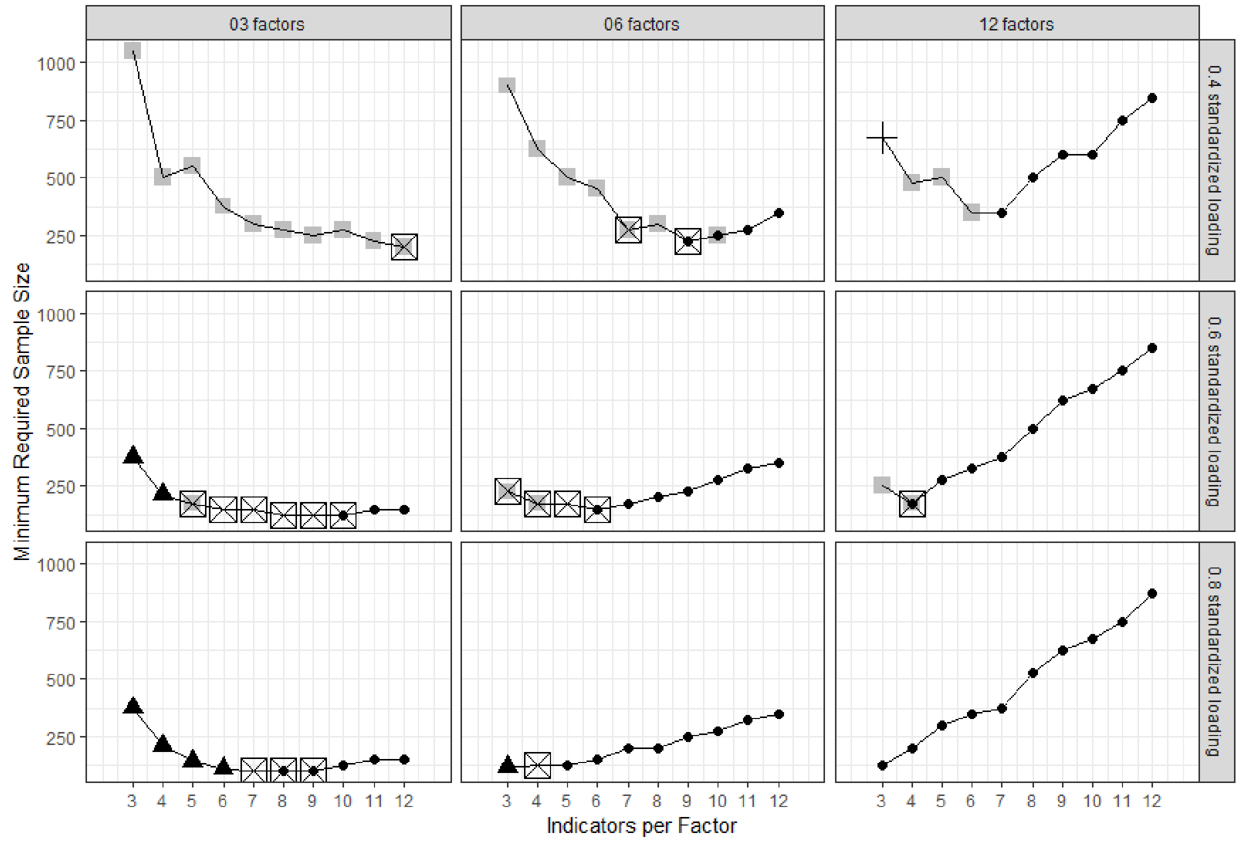

Figure 1. Minimum required sample size by number of indicators per factor for nine confirmatory factor analysis models with varying numbers of factors and standardized loadings.

adequate percentage of proper solutions

RMSEA – adequate power for the RMSEA test of close fit

RBse – reasonable average relative bias in the standard error of the parameter estimate

power – adequate power for the parameter of interest

ChiSq – reasonable rejection rate for the model chi-square statistic

RBse & power

power & ChiSq

RBse & ChiSq

RBse & power & ChiSq