

Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General

Theodoros A. Kyriazos

Department of Psychology, Panteion University, Athens, Greece
Email: th.kyriazos@gmail.com

How to cite this paper: Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology*, 9, 2207-2230. <https://doi.org/10.4236/psych.2018.98126>

Received: July 26, 2018

Accepted: August 21, 2018

Published: August 24, 2018

Copyright © 2018 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Adequate statistical power contributes to observing true relationships in a dataset. With a thoughtful power analysis, the adequate but not excessive sample could be detected. Therefore, this paper reviews the issue of what sample size and sample power the researcher should have in the EFA, CFA, and SEM study. Statistical power is the estimation of the sample size that is appropriate for an analysis. In any study, four parameters related to power analysis are Alpha, Beta, statistical power and Effect size. They are prerequisites for a priori sample size determination. Scale development in general and Factor Analysis (EFA, CFA) and SEM are large sample size methods because sample affects precision and replicability of the results. However, the existing literature provides limited and sometimes conflicting guidance on this issue. Generally, for EFA the stronger the data, the smaller the sample can be for an accurate analysis. In CFA and SEM parameter estimates, chi-square tests and goodness of fit indices are equally sensitive to sample size. So the statistical power and precision of CFA/SEM parameter estimates are also influenced by sample size. In this work after reviewing existing sample power analysis rules along with more elaborated methods (like Monte Carlo simulation), we conclude with suggestions for small samples in factor analysis found in literature.

Keywords

Sample Size, Sample Power, SEM, CFA, EFA, Psychometrics, Monte Carlo Simulation, Test Development

1. Introduction

An adequate sample size or more precisely sample power is of primary concern when designing a study (Tabachnick & Fidell, 2013). Adequate statistical power

contributes to observing true relationships in the dataset (Wolf, Harrington, Clark, & Miller, 2013). Therefore, this paper considers the following question: *what sample size should the researcher acquire in three different study designs?* 1) Exploratory Factor Analysis (EFA); 2) Confirmatory Factor Analysis (CFA); 3) Structural Equation Modeling (SEM).

Estimation of the power of a statistical analysis during the planning of the study is generally accepted as a good practice (Thomas, 1997; Schumacker & Lomax, 2015). During prospective power analysis, the researcher estimates the minimum required sample size to achieve the maximum level of statistical power for a hypothesized effect size under a specified statistical significance level (Wilcox, 2008 cited in Wang, Watts, Anderson, Little, 2013). Thus, the sample size has an impact on the precision of all statistical estimates, including those made in EFA (Thompson, 2004). Specifically, in EFA the replicability of a factor structure is partially dependent on the sample size of the initial analysis. As a rule, the factor pattern developed by a large-scale factor analysis is probably more stable than that based on a small sample size (DeVellis, 2017). The bottom line question is “How large is large enough?” (Kline, 2016) and there is no easy answer to it because like many other statistical procedures both the number of variables analyzed and the absolute number of subjects should be taken into account (DeVellis, 2017), in addition to other issues indicating if the data is “strong”. As a general rule, the stronger the data, the smaller the required sample to achieve adequate accuracy (Costello & Osborne, 2005). “Strong data” in factor analysis is indicated by high communalities, no cross-loadings, strong primary loadings per factor and also additional variables like the nature of the data, number of factors, number of items per factor (MacCallum, Widaman, Zhang, & Hong, 1999; Fabrigar et al., 1999; Costello & Osborne, 2005; DeVellis, 2017). In practice, these conditions are very difficult to be simultaneously true (Mulaik, 1990; Widaman, 1993; Costello & Osborne, 2005).

On the other hand, SEM is used most often to confirm a prior hypothesis, in contrast to the exploratory nature of factor analysis thus, planning is crucial for any SEM analysis (Tabachnick & Fidell, 2013), including CFA. SEM is also a large sample approach (Kline, 2016). It is generally accepted that problems may arise due to a small sample size. Some of them include—but they are not limited—to estimation convergence failure, improper solutions (e.g., Heywood cases), inaccurate parameter estimates and model fit statistics (Wang & Wang, 2012). Additionally, SEM flexibility allowing the examination of complex associations, multiple data types, model and/or group comparisons thus, developing general rules regarding sample size requirements are impractical (MacCallum et al., 1999; Wolf et al., 2013). In CFA, being a SEM category, sample size depends on a number of features like study design (e.g. cross-sectional vs. longitudinal); the number of relationships among indicators; indicator reliability, the data scaling (e.g., categorical versus continuous) and the estimator type (e.g., ML, robust ML etc.), the missing data level and pattern and model complexity (Brown, 2015). Thus, determining sample size is approximated by power analysis

(Brown, 2015; Kline, 2016; Byrne, 2012; Wang & Wang 2012).

The research questions answered in the next sections are as follows: 1) What is power analysis? 2) Why does sample power need to be taken into account in factor analysis? 3) What power analysis methods exist in CFA and SEM framework? 4) What can the researcher do when the sample size is small?

2. Power Analysis Basics

Statistical power is the estimation of the sample size that is appropriate for an analysis (Cohen, 1988, 1990, 1992). The statistical power of a study is the likelihood of detecting an actually present effect (Coolican, 2014). It could be compared to the precision power of a microscope in the laboratory. If using a low-magnification microscope fine details are hard to detect. In a similar way in a study of low power, more fine effects could be missed out (Barker, Pistrang, & Elliott, 2016).

In any study, there are four parameters related to power analysis as reviewed by Barker, Pistrang & Elliott (2016): 1) The size of the sample (N). 2) The probability of identifying a non-existing effect is called *Alpha* (α). This kind of error has termed Type I error (or false positive). In most psychological research, alpha is set by arbitrary convention at .05 (see also Wolf et al., 2013). 3) The probability of not identifying an existing effect is called *Beta* (β). This is the Type II error (or false negative). The probability to identify an effect that really exists is calculated by subtracting beta from one ($1 - \beta$) and the result is defined as *statistical power* (Cohen, 1988). The desired level of statistical power is .80 (Cohen, 1988, 1992) and a minimum is .50, i.e. a 50% probability to detect an existing effect. 4) *Effect size* is a measure of the strength of the examined relationship. Effect sizes are described as small, medium, and large and are different for each statistical test (Barker, et al., 2016). The statistical power is best considered during study planning to determine the appropriate sample size (Tabachnick & Fidell, 2013; Thomas, 1997; Wilcox, 2008). The four above estimates are prerequisite for a priori sample size determination (see Table 1). Omitting this step during the planning stage could potentially mean failure to detect a significant effect (Tabachnick & Fidell, 2013).

Table 1. Prerequisites for a priori sample size determination (Cohen, 1988).

Parameter	Level	Error type	Description
<i>Alpha</i>	.05	Type I error or False positive	The probability of identifying a non-existing effect
<i>Beta</i>	-	Type II error or False negative	The probability of not identifying an existing effect
<i>Statistical power</i> ($1 - \beta$)	min > .50 Ideally > .80		The probability to identify an effect that really exists
<i>Effect size</i>	Small, Medium, or Large		The strength of the examined relationship

A question emerging is “Then, why not obtaining huge sample power?” following the rule of thumb suggesting that the larger the sample, the better (Thompson, 2004; p. 24). Cohen (1990) noted that unduly large samples, beyond what is required to achieve statistical power are a waste of research effort, and could overstate unimportant effects (Barker et al., 2016). Thus, an equilibrium is sought between a too small sample size that could fail to uncover crucial effects and a too large sample adding extra cost and time to the study (Wang, Watts, Anderson, & Little, 2013; Nicolaou & Masoner, 2013). With a thoughtful power analysis, the adequate but not excessive sample could be detected (Du, Zhang, & Yuan, 2017). Instead, when the luxury of a large sample is available, a better research strategy is suggested: to implement multiple smaller studies on different populations (Barker, et al., 2016).

Finally, statistical power and sample size can be estimated with different methodologies before data collection (See Table 2). This type of analysis is called *a priori* or *prospective* power analysis, whereas if this analysis is carried out after data collection is called *post-hoc* or *retrospective* (Wang et al., 2013). Figure 1 contains common myths (fallacies) about sample power related to retrospective power analysis by Wang, Watts, Anderson, and Little (2013).

Table 2. Statistical power and sample size estimation methodologies based on their time of implementation.

Method	Description	Usage
<i>a priori</i> or <i>prospective</i>	Sample power is estimated before data collection	estimating the minimum sample size required for a power level is the recommended course of action
<i>a posteriori</i> or <i>retrospective</i> ,	Sample power is estimated after data collection. Also termed <i>post hoc</i> or <i>observed</i>	A controversy field due to misuse in applied research

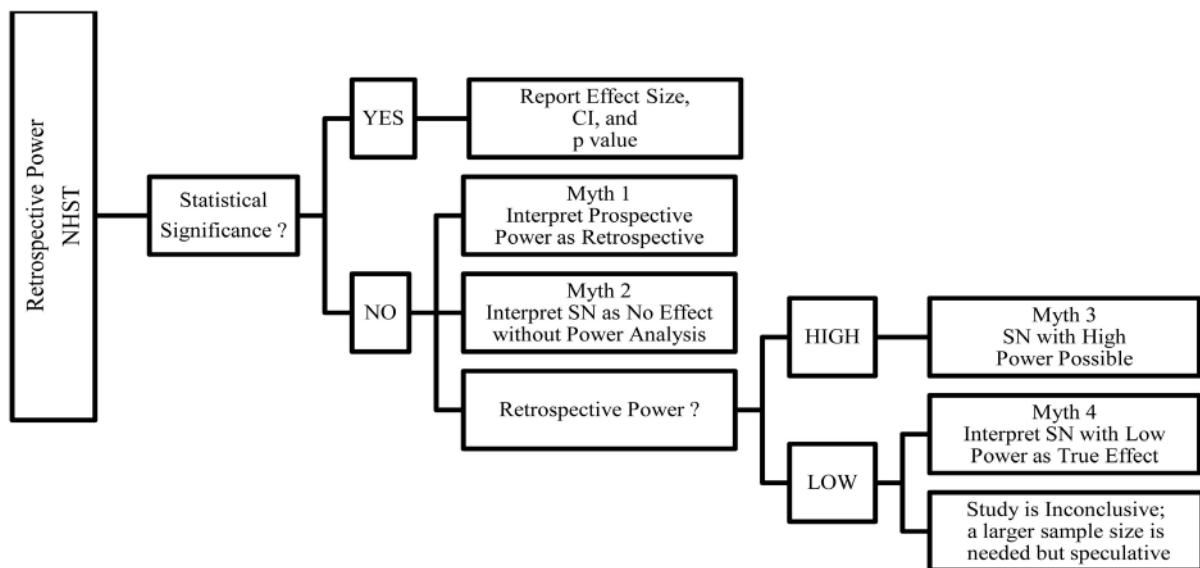


Figure 1. Common myths related to retrospective power analysis as postulated by Wang, Watts, Anderson, and Little (2013). Note: This figure is based on a flow chart by Wang, Watts, Anderson, and Little (2013, page 738), NHST = Null Hypothesis Significance Testing, CI = Confidence Interval, SN = Statistical Non-significance.

3. Sample Power Implications for Factor Analysis

Like in inference statistics, in factor analysis too, it is considered a good practice to a priori determine the minimum sample size required to achieve an acceptable level of statistical power for the factor structure under evaluation (Thomas, 1997; Schumacker & Lomax, 2015; McQuitty, 2004; Singh et al., 2016). Scale development in general and factor analysis in particular, are large sample size methods (DeVellis, 2017; Costello & Osborne, 2005 to quote a few). This requirement becomes more crucial when SEM (more precisely CFA) is used as a validation method because SEM is also a large sample method (Kline, 2016; Brown, 2015; Shumacker & Lomax, 2016; Wang et al., 2013; Wang & Wang, 2012). However, the existing literature as Brown (2015: p. 380) comments “*provides little guidance on this issue*”.

Generally, in Factor Analysis (FA) sample size is considered a top priority issue (Comrey & Lee, 1992; Costello & Osborne, 2005; Gorsuch, 1983; Shumacker & Lomax, 2012) because FA is a method essentially based on correlation coefficients. Whether the coefficient is an adequate estimate of the population correlation taps statistical inference and validity, i.e. the more stable the sample correlations, the more valid the scores (Schumacker & Lomax, 2015; Finch, French, & Immekus, 2016; Tabachnick & Fidell, 2013). On the contrary, smaller samples potentially produce unstable correlation estimates, more prone to outliers (Finch et al., 2016).

Additionally, besides validity, the sample size has also an impact on reliability because the more reliable the scale the lower the required sample size to achieve the desired statistical power for a specific test as DeVellis (2017) explains. DeVellis gives an illustrative example for his argument: for $N = 50$ if two scales have a reliability of .38 and they are correlated with $r = .24$ at a significance level of $p < .10$. If the reliability of the measure employed is increased at .90 the significance level becomes $p < .01$. If reliability remains at .38, twice as many participants would be needed for the correlation to reach $p < .01$ level. Other parameters affecting the sample size in FA is the number of factors and the number of items present (DeVellis, 2017). More details about how sample size can affect EFA and CFA research follow (and SEM more generally).

3.1. EFA Sample Size Considerations

Generally, in a large sample correlations estimates are regarded as more reliable than in a small sample. Other EFA parameters crucial for the sample size is the magnitude of population correlations and number of factors of the estimated solution. The strongest the correlations and the fewer the factors the smaller the required sample (Tabachnick & Fidell, 2013). Therefore, the sample size is by and large specified by the nature of the data (Fabrigar et al., 1999). The stronger the data, the smaller the sample can be for an accurate analysis and “strong data” within the EFA framework means high communalities and absence of cross-loadings and strong primary factor loading on the intended factor (Costello &

Osborne, 2005; Thompson, 2004). In empirical research, however, these conditions are hard to find (Mulaik, 1990; as quoted by Costello & Osborne, 2005).

In a similar vein, the Monte Carlo simulation work by Guadagnoli and Velicer (1988) suggested that the crucial parameter in EFA sample size is the degree of factor saturation by the measured variables. Guadagnoli and Velicer (1988) focused on the factor pattern stability as a function of the population pattern for: 1) a range of sample sizes (for $N = 50, 100, 150, 200, 300, 500,$ and 1000); 2) a range of measured variables (for $p = 36 - 144$); 3) a range of structure coefficients (for $a = .40, .60,$ and $.80$); and 4) range of numbers of factors (for $m = 3, 6,$ and 9) as reproduced by Dimitrov (2012). They proposed that factor replicability is more likely when: 1) factors have at least four measured variables with structure coefficients $> |.6|$, irrespectively of the size of the sample; 2) for $N > 150$ factors are defined with 10 or more structure coefficients of about $|.4|$ (and low p/m ratio), when $300 \leq N \leq 400$ (Guadagnoli & Velicer, 1988: p. 274 quoted in Dimitrov, 2012). Additionally, replicability of the factor pattern was also achieved when: 4) $a = .80$ across all conditions (as reviewed by Dimitrov, 2012; Thompson, 2004). See also Figure 2 about EFA sample size basics.

3.2. CFA and SEM Sample Size Considerations

SEM is a method that is estimated based on covariances. Covariances, like correlations, turn out to be unstable if assessed over small samples. Generally, the findings of Velicer and Fava (1998) see also Guadagnoli & Velicer, 1988) about the size of the factor loadings and the number of variables as a function of the sample size are important elements for obtaining a good CFA or SEM model as

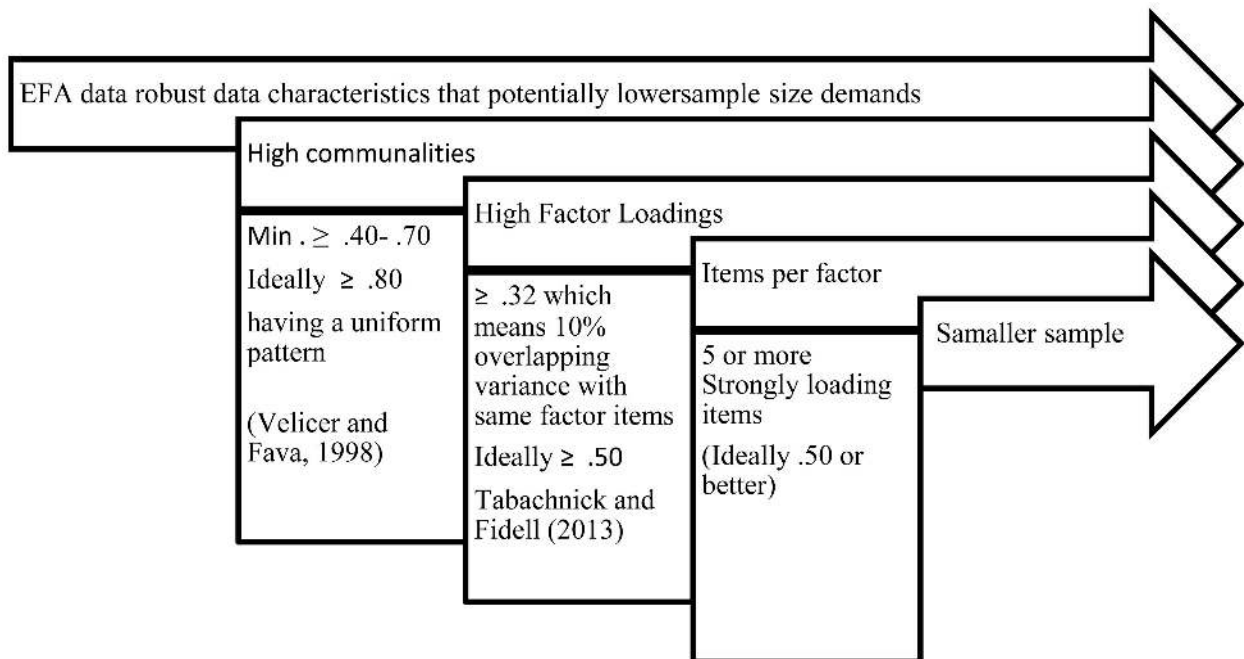


Figure 2. EFA indicators of strong data that potentially may require smaller sample size because the stronger the data the smaller the sample size (Costello & Osborne, 2005).

well. Moreover, parameter estimates, chi-square tests and general goodness of fit indices are equally sensitive to sample size. This means—with a risk of oversimplification—that as a rule models having robust parameter estimates and variables with high reliability may require smaller samples in CFA and SEM too (Tabachnick & Fidell, 2013). SEM is a large-sample technique (Kline, 2016) for the reasons described next.

First, the statistical power and precision of a CFA (and SEM in general) model parameter estimates are influenced by the sample size (Brown, 2015). During a CFA a hypothetical model is tested. When the data do not fit the hypothesized model, we modify the model to improve fit, generally based on modification indices. This hypothesis testing involves statistical power considerations. However, in CFA, power is redefined as the ability to retain the null hypothesis and reject the alternative hypothesis. However, determining the sample power and/or sample size for a CFA analysis is more complicated in comparison to EFA because CFA models are based on theoretical models potentially having numerous parameter estimates dependent as a rule to each other adding up parameters affecting latent variables (like covariances and standard errors) that become less accurate in small samples (Kline, 2016).

Apart from that, CFA requires model comparison, even comparison of nested models in a single dataset. The power for this hypothesis testing depends on the true population model, the level of significance and degrees of freedom of the model as well as on the sample size which in turn requires determining an effect size and alpha level of significance. However, a sample size is determined given power, effect size, and alpha (Schumacker & Lomax, 2015).

Moreover, particular fit indices “react” differently in small sample sizes along with model estimators, model complexity, multivariate normality assumption and variable independence (Fan & Sivo, 2007; Saris, Satorra, & van der Veld, 2009 as cited in Byrne, 2012). The chi-square test is perhaps the most notoriously sensitive fit measure to sample size (Kline, 2016; Finch, et al., 2016). In small sample sizes (<200) the chi-square may fail to reject an unfitting model while in a large sample may falsely reject an adequate model (Gatignon 2010; Singh et al., 2016). This happens because the chi-square test equals $(N - 1) F_{min}$ and this value is significant when the model fit is inadequate and the sample size is large (as described in Byrne, 2012 and Jöreskog & Sörbom, 1993). However, large samples are crucial for models with accurate parameter estimates, especially when the assumption of normality is rejected (Byrne, 2012 also quoting MacCallum et al., 1996). Therefore, the chi-square to the degrees of freedom ratio (chi-square/df) was introduced instead (Wheaton, Muthén, Alwin, & Summers, 1977; Jöreskog & Sörbom, 1993) as Brown (2015) comments. However, the chi-square/df ratio is just as sensitive to sample size as chi-square (Brown, 2015; also quoting Wheaton, 1987). Nevertheless, current reporting ethics use it, so it would be an omission not to report it. However, it is usually reported along with other fit measures to minimize this oversensitivity to sample size.

The Root-mean-square error of approximation (RMSEA; ϵ) is relatively in-

sensitive to sample size (Brown, 2015). However, Hu and Bentler (1999) note that with a small sample size, RMSEA is oversensitive in rejecting true population models (Byrne, 2012). Additionally, the width of RMSEA confidence intervals is affected by sample size and model complexity (MacCallum et al., 1996; Brown, 2015; Byrne, 2012). For a small N and a large number of estimated parameters (a complex model), the confidence intervals will be wide (Byrne, 2012; Brown, 2015). On the other hand, for moderate N s and low complexity models, obtaining a narrow confidence interval is more likely (MacCallum et al., 1996 cited in Byrne, 2012). In a Monte Carlo study by Curran et al. (2002) was reported that when N was >200 the RMSEA was accurate for models with moderate misspecifications. MacCallum and Hong (1997) also propose that RMSEA is more efficient than the GFI and AGFI for power analysis (Loehlin & Beaujean, 2017). Other fit indices are also affected by sample size. Specifically, TLI like RMSEA is prone to false model rejections when the sample size is not adequate (Hu & Bentler, 1999 cited in Brown, 2015). Finally, the CFit test is adversely affected by small sample size like any other test of significance (Brown, 2015).

Except for model fit indices sample size also has an impact on the model estimated parameters, the method of estimation, the extent of harmless model misspecification, data normality (see also Table 3). Finally, the size of standardized residuals is a function of the size of the sample (Brown, 2015). As a rule, larger samples are related to larger standardized residuals. This happens because

Table 3. Factors affecting sample size requirements in SEM and CFA (Kline, 2016: p. 15).

SEM and CFA
Model Complexity or/and number of model parameters estimated
Analyses in which all outcome variables are continuous
Normally distributed data, and there are no
Linear effects existing in data
Existing interactions between data
Estimation method
The lower the reliability of the scores the higher the required sample size
Is it a latent variable models or observed variable model?
Less precise data requires larger samples
Missing data require larger sample sizes
CFA in particular
Low number of indicators for the constructs of interest per factor requires larger samples
Lower number of indicators per factor requires larger samples
Indicators that covary highly with multiple factors require larger samples
If the number of factors is high a larger sample is needed
If covariances between factors are low a larger sample is needed

for the fitted residuals the size of their standard errors is frequently inversely associated to sample size. Thus, the interpretation of the standardized residuals should be made with the sample size in mind. Modification indices are equally affected by sample size, proposing parameter additions with an unsubstantial magnitude when the sample size is large. On the other hand, a small sample size (e.g. $N = 100$; Silvia & MacCallum, 1988) may cause specification searches suggesting incorrect model revisions. Thus, as CFA is a large sample method, minor effects are sometimes falsely proposed to have statistical significance. When working with large samples, it is important, as Brown consults, to demonstrate that the parameter estimates have a substantively meaningful magnitude (Brown, 2015: p. 115).

Additionally, with a small sample size, technical problems are more likely too. Inadmissible CFA solutions may include Heywood cases, i.e. negative variance estimates or estimated absolute correlations > 1.0 . Experts warn that small samples ($N < 100 - 150$) and few indicators per factor (< 3) are more prone to non-convergence or improper solutions (Kline, 2016 also quoting Marsh & Hau, 1999). Generally, if the sample size is small more observed indicators per factor could alleviate its impact (Marsh et al., 1998; Marsh & Hau, 1999). Correspondingly, if the sample is large could yield robust factors even with few indicators per factor. E.g. a CFA model with 6 - 12 indicator variables per factor could be specified with $N = 50$, while $N > 100$ would be necessary for a CFA model with 3 - 4 indicators per factor (Boomsma, 1985; Marsh & Hau, 1999). Finally, a CFA model with 2 indicators per factor $N > 400$ would be necessary (Marsh & Hau, 1999; Boomsma & Hoogland, 2001). Besides ML is notorious for non-convergence and small samples are a possible cause (Finch et al., 2016). However, Wang and Wang (2012) comment that a factor structure with a large number of indicators per factor, it is often difficult to be validated because numerous error terms will be possibly correlated.

Finally, some aspects/categories in CFA potentially affected by sample size also include: 1) Measurement invariance; 2) Item parceling. In measurement invariance, researchers use the $\Delta\chi^2$ criterion to compare the fit of nested models (see Cheung & Rensvold, 2002). This criterion is equally sensitive to sample size to the chi-square (Byrne, 2012; Brown, 2015). Additionally, the effects of using item parcels can differentiate with sample size (Hau & Marsh, 2004). This sample size may be a crucial parameter when deciding whether to use item parceling or not (Byrne, 2012). Furthermore, the evaluation of CFA sample size must be made in regard to its suitability for ML estimation method because if ML is not possible alternative analytic approaches or estimators (e.g., robust ML) could be used (Brown, 2015). With these new robust estimators, the need for a large sample is less imperative (Raykov, 2012) because under certain conditions can handle as few as 60 participants (see Bentler & Yuan, 1999; Wolf et al., 2013; Chumney, 2013) irrespectively of the normality assumption (Wang & Wang, 2012; Brown, 2015; Kline, 2016). See also **Figure 3**.

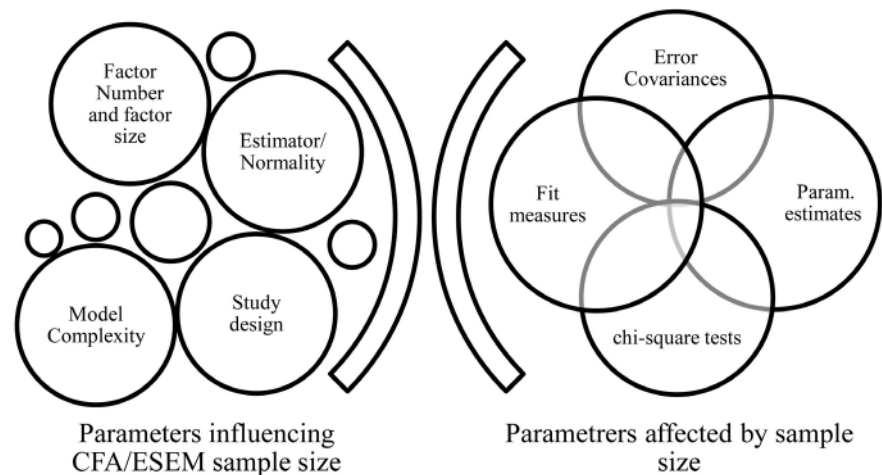


Figure 3. CFA/SEM parameters influencing sample size and parameters affected by sample size.

4. Sample Power Analysis Rules

These traditional rules of thumb about sample size along are summarized next.

4.1. Rules of Thumb

Minimum sample sizes in absolute N s were the first rules of thumb, suggesting that any $N > 200$ offers adequate statistical power for data analysis (Hoe, 2008; Singh et al., 2016). The same N is also proposed by Comrey (1988) as generally adequate for a measure having up to 40 items. A sample of 300 cases has also been suggested (Tabachnick & Fidell, 2013). Comrey and Lee (1992); and Comrey et al., 1973) graded a factor analysis sample of 50 as very poor, 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 as excellent (quoted also by Costello & Osborne, 2005; DeVellis, 2017; Williams et al., 2010 and others). According to Kline (2016) though it is difficult to set a minimum sample size in SEM studies a median sample based on study reviews is $N = 200$ (MacCallum & Austin, 2000). However, he adds that $N = 200$ may be too low for complex models with non-normal distributions with missing data. He also comments that N s < 100 , as a rule, generate untenable results. Finally, for a multi-group CFA, a general rule of thumb is 100 participants in each group (Kline, 2016; Wang & Wang, 2012).

Over the years, rules of thumb (or so-called blue-chips, Nicolaou & Masoner, 2013) proposed that the ratio of the number of people (N) to the number of measured variables (p) must be considered. Based on these assumptions, sample size should be greater than the number of variables i.e. $N > p$ (Nunnally & Bernstein, 1994 as quoted in Dimitrov, 2012). The recommended $N:p$ ratios became progressively larger, ranging from 5 with a minimum $N > 100$ (Gorsuch, 1983; cited in Dimitrov, 2012), to 10 (Nunnally & Bernstein, 1967; Everitt, 1975). A widely accepted ratio is 10 cases per indicator variable (Nunnally & Bernstein, 1967 quoted by Wang & Wang, 2012). Tinsley and Tinsley (1987) suggested a

ratio of 5 to 10 participants per item for $N = 300$ noting that for $N > 300$ this ratio can become progressively lower (as noted by [Devellis, 2017](#)). For scale development, a general rule is that for a unidimensional scale constructed out of a 20-items pool a $N = 300$ could be sufficient ([DeVellis, 2017](#)). Likewise, this ratio for “traditional multivariate statistics” can be 20 cases per measured variable ([Shumacker & Lomax, 2016: p. 240](#)) in line with a similar rule of thumb used in linear regression ([Lomax & Hahs-Vaughn, 2013](#)) but in SEM this can get as high as 100 - 500 or more subjects per study ([Shumacker & Lomax, 2016: p. 240](#)).

Another variation of the N:p rule pertinent in CFA/SEM is the N:q rule, i.e. the number of cases (N) to the number of estimated parameters (q). This rule taps the model precision, i.e. the ability of the parameter estimates to approximate true population values. Model precision is also a function of the bias of the parameter estimates and their standard errors ([Brown, 2015](#)). This ratio for CFA can range from 5 to 10 cases ([Bentler & Chou, 1987](#); [Bollen, 1989](#)). If the data is highly kurtotic an $N:q > 10$ was proposed ([Wang & Wang, 2012](#) quoting [Hoogland & Boomsma, 1998](#)). On the other hand, even for latent variable models with continuous outcomes and normal distribution using ML [Jackson \(2003\)](#) suggested a sample-size to parameters ratio of 20:1 or at least 10:1. Results with lower ratios are progressively less trustworthy and the risk of technical problems looms larger (see more details on [Kline, 2016](#)).

However, strict rules on sample size have mostly disappeared ([Costello & Osborne, 2005](#)). Instead, new rules based on a number of Monte Carlo simulation studies gradually emerged.

4.2. Rules Based on Monte Carlo Simulation Studies

Monte Carlo methods are mathematical methods using random sampling and computer simulation to solve problems ([Wang & Wang, 2012](#)) under different CFA/SEM conditions and different N s is one of them i.e. statistical power ([Brown, 2015](#)).

Findings suggest (see also [Table 4](#)) that SEM models could be safely evaluated with small samples ([Hoyle, 1999](#); [Hoyle & Kenny, 1999](#); [Marsh & Hau, 1999](#)), but generally $N = 100 - 150$ is set as a minimum sample size for SEM research ([Anderson & Gerbing, 1988](#); [Ding, Velicer, & Harlow, 1995](#)) while others set this minimum to $N = 200$ ([Hoogland & Boomsma, 1998](#); [Boomsma & Hoogland, 2001](#)) as per [Loehlin \(2004\)](#). In a similar vein, [Kelloway \(2015\)](#) commented that [Anderson and Gerbing \(1984\)](#) also used a Monte Carlo simulation reaching to a similar conclusion, i.e. that small samples in CFA ($N < 100$), caused convergence failures and improper solutions in models with < 2 indicators per latent variable. The use of 3 indicators per latent variable along with $N > 200$ led to almost zero convergence failures and no improper solutions.

[MacCallum, Widaman, Zhang, and Hong \(1999\)](#) in a very influential study on sample size in factor analysis also suggested that 100 - 200 cases are adequate when: 1) multiple indicators define a factor; 2) marker variables have loadings $> .7$ and 3) communalities are about .5 (ideally $> .6$ or $> .7$ on average). Low

Table 4. Selected results from monte carlo simulation studies.

Estimator	Sample size Recommendation	Studies recommending it
ML with multivariate normal data (continuous)	1) 100	1) Anderson & Gerbing (1984)
	2) 200-400	2) Jackson (2001)
	3) N:q = 5:1	3) Tanaka (1987)
	4) N:q = 10:1	4) Bentler & Chou (1987)
	5) 30-460	5) Wolf et al. (2013)
MLM	≥250	Hu & Bentler (1999); Yu & Muthén (2002)
Bootstrap with nonnormal continuous data	≥200 - 1000	Nevitt & Hancock (2001)
MLR with continuous nonnormal data with missing values	>400	Yuan & Bentler (2000)
Robust DWLS/ WLSMV with binary or ordinal data	≥200 - 500	Bandalos (2014) Forero, Maydeu-Olivares, & Gallardo-Pujol (2009)
MLR for binary and ordinal variables	≥200 - 500	Bandalos (2014)

Note. This table is mainly based on a Table by Newsom (2018: p. 1).

communalities, a small number of weakly determined factors with 3-4 indicators per factor increase the required sample to 300 cases and when all conditions are adverse, i.e. communalities are low, there are many weakly determined factors the cases required is 500 (Tabachnick & Fidell, 2013; Thompson, 2004; Dimitrov, 2012). In a nutshell, MacCallum et al. (1999) proved that model parameters including (but not limited) to communalities, and factor determinacy can affect the accuracy of the parameter estimates and model fit statistics as a function of sample size.

Muthén and Muthén (2002) concluded that for a CFA model with three factors and five continuous indicators per factor to reach a power of .81 in rejecting the hypothesis that the factor correlation is zero, the required sample size was: 1) $N = 150$ for normal indicators with no missing values, 2) $N = 175$ for normal indicators having missing values, 3) $N = 265$ for non-normal indicators and no missing values, and 4) $N = 315$ for non-normal indicators having missing values (Dimitrov, 2012).

Regarding the impact of factor strength as demonstrated by the magnitude of regressive effects of a model on sample size, Wolf et al. (2013) in their Monte Carlo simulation study reported that both very weak and very strong effects may demand larger samples, and this effect is more evident in weak magnitude factors (Wolf et al., 2013). These findings (see also Figure 4) actually question both the “one size fits all” and the rules of thumb approach to CFA and SEM research, as noted by Wolf et al. (2013). On the other hand, Monte Carlo simulation studies results were questioned as having a limited generalizability (Brown, 2015). More model-based sample power methods of determining sample size and sample power are described next.

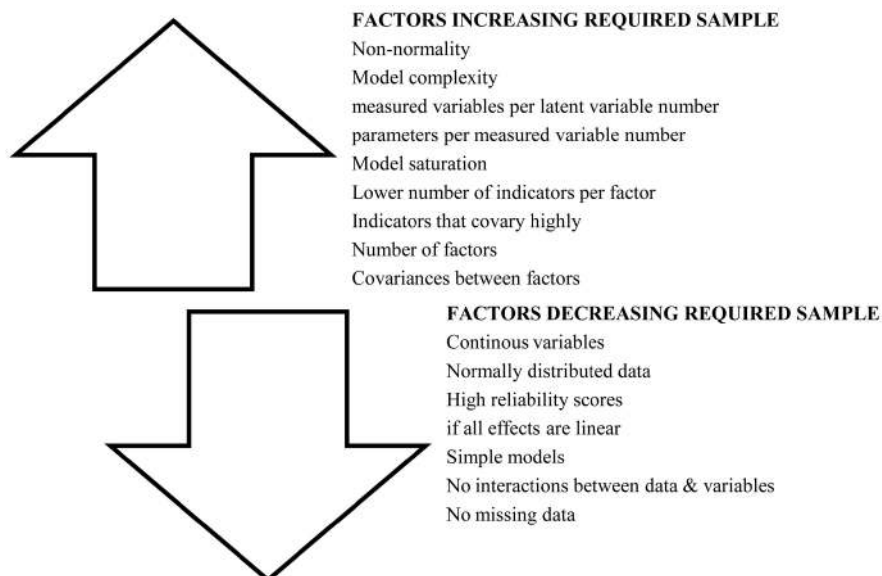


Figure 4. Factors that potentially increase and decrease sample size in CFA/SEM (Kline, 20016; Nicolaou & Masoner, 2013).

5. Sample Power Analysis Methods

Instead of rules of thumb, sample size and power are suggested to be determined considering models, data and empirical context (Brown, 2015; Wang & Wang, 2012). Generally speaking, the power in an inferential statistics test is the probability that one will reject the hypothesis tested if it is false. In CFA and SEM four things are required to determine the power of a test: 1) a model, 2) an alternative model to be compared to the first one, 3) the targeted level of significance, 4) the sample size N (Loehlin, 2004; Schumacker & Lomax, 2015). Based on these elements the methods described next calculated the adequate sample size in CFA and SEM models.

5.1. The Critical N (CN) Statistic

Hoelter (1983) introduced the Critical N (CN) statistic for the evaluation of SEM sample size, where $CN \geq 200$ was considered adequate. Based on the model degrees of freedom a critical chi-square value is calculated. CN proposes the sample size at which the F_{min} value rejects H_0 (Schumacker & Lomax, 2015 also quoting Bollen & Liang, 1988; Bollen, 1989). After data collection and SEM model specification, we could estimate the post-hoc sample power with the non-centrality parameter (NCP or λ). Sample size N equals $(NCP/F_{min}) + g$. Hence, we could a-priori obtain the F_{min} value from our model, calculate the NCP for a given df, critical chi-square and power then calculate the sample size (N) using these values. McDonald and Marsh (1990) studied non-centrality and model-fit issue further by evaluating how nine fit indices perform with regards to non-centrality and sample size. For further details, refer to Schumacker and Lomax (2015) who are the source of this paragraph.

5.2. The MacCallum et al. (1996) Not-Close Fit Method

MacCallum, Browne, and Sugawara (1996) suggested a different approach to testing model fit using power and the root-mean-square error of approximation (RMSEA; ϵ). They introduced the RMSEA confidence intervals, rather than a single point suggesting null and alternative RMSEA but researchers can also define their own. This approach evaluates power, given exact fit (H_0) where RMSEA is zero, close fit (H_0) where $RMSEA \leq .05$ and not close fit (H_0) where $RMSEA \geq .05$. They also offered a SAS code for calculating power for a given sample size or sample size for a given power using RMSEA for an exact fit, for a close fit, and for not a close fit. They proposed that an RMSEA value of .05 - .08 is satisfactory along with other fit measures, and a power of .768. Power is defined as the probability of not rejecting the null hypothesis, therefore a close fit of the sample covariance matrix with the model-implied covariance matrix (Schumacker & Lomax, 2015; Loehlin & Beaujean, 2017).

MacCallum, Lee, and Browne (2010) further elaborated on sample power in CFA and SEM. Hancock and French (2013) discussed the use of the non-centrality parameter (NCP; λ) and root-mean-square error of approximation (RMSEA; ϵ) when testing the null and alternative CFA/SEM models. See Schumacker & Lomax (2015) for more details.

5.3. The Satorra Sarris Method (1985)

Satorra and Saris (1985) and Saris and Satorra (1993) introduced an alternative approach for evaluating a CFA/SEM model power (Schumacker & Lomax, 2015).

The method is based on the idea that a moderately misspecified model fit test statistic follows a non-central chi-square distribution. The chi-square of the misspecified model approximates the non-centrality parameter (NCP or λ) of the non-central chi-square distribution. NCP is estimated as $\chi^2 - df_{\text{model}}$ according to the weighted least squares estimation. Once the NCP parameter is calculated, statistical power is obtained either from a table for non-central chi-square distribution for given degrees of freedom and a level (Saris & Stokhorst, 1984) or calculated by statistical packages (Wang & Wang, 2012; Schumacker & Lomax, 2015). The application of the method to estimate statistical power and derive sample size requires a sequence of five steps (Brown, 2015; Wang & Wang, 2012).

In an attempt to compare the Satorra and Saris method (1985) with the MacCallum et al. method (1996), Lee, Cai, and MacCallum (2012) remarked that in the former misspecification of particular parameters and their magnitudes is required as an input. In the later, the misfit of the hypothesized model or fit difference is required. Thus, when data is not enough or parameter values are unreliable, e.g. on research inception, then the latter approach could be more appropriate demanding substantially fewer user data (Lee, Cai, & MacCallum, 2012). A drawback of this approach is that it must be repeated for every individual parameter for which an estimate of power is desired (Kline, 2016). See also **Table 5** for Method steps.

Table 5. Satorra and Saris (1985) method steps for calculating sample size.

Step	Description
Step 1	Model specification with hypothesized population parameter values and employ a null covariance matrix (i.e., a matrix with 1s on the diagonal and 0s off the diagonal)
Step 2	Check of accuracy. The H1 model is freely estimated with the fitted covariance matrix from step 1 as input. If the estimated parameters estimated match those in Step 1, then we can proceed to the next step.
Step 3	Specification of H0. Select a sample size and specify a misspecified model by restricting the targeted parameter to zero (or the value expected under the null hypothesis), and then run the model using the generated covariance matrix as data input.
Step 4	Use the model chi-square from Step 3 as an approximate NCP to compute statistical power of detecting the effect of interest at a given level
Step 5	Repeat Steps 3 and 4 with various sample sizes and compute corresponding power values. The sample size corresponding to a power > 0.80 is an estimation of the required sample size.

Note. Steps are from Wang & Wang (2012) and Brown (2015, p. 385).

5.4. The Monte Carlo Approach

Muthén and Muthén (2002) demonstrated how the CFA/SEM sample power can be a priori determined with Monte Carlo simulation (Loehlin & Beaujean, 2017).

Monte Carlo simulation estimates the proportion of generated samples where the null hypothesis is correctly rejected (Bandalos & Leite, 2013; Kline, 2016). To estimate power and sample size for a model with Monte Carlo simulation a hypothesized population value for each model parameter is defined based on theoretical or empirical findings. Then a large number of samples are randomly generated. The model is estimated in each of the generated samples (Wang & Wang, 2012). Then the results of all samples are averaged (parameter values, standard errors, fit statistics). Based on these averaged results precision and power of the estimates are examined (i.e., the percentage of samples in which the parameter significantly differs from zero). Various sample sizes are examined to find out the required N to achieve parameter estimates with the desired power and precision. The analysis will proceed by examining larger sample sizes (and other seed values), to achieve stability once a suitable N has been identified. This is accomplished by changing the number of observations (Brown, 2015). The criteria suggested by Muthén and Muthén (2002) for sample size calculation is the following: 1) parameter and standard error bias < 10% for each model parameter; 2) standard errors bias < 5% for parameters that the power analysis targets and 3) coverage ranging from .91 to .98. The required sample size is specified when the power of salient model parameters is $\geq .80$ (Cohen, 1988; Brown, 2015; Dimitrov, 2012). The Monte Carlo simulators available can be programmed to reproduce a specific amount of non-normality and missing data. Nonetheless, they do not handle joint skewness and kurtosis of the distribution, i.e. multivariate non-normality (Brown, 2015).

5.5. Kim's (2005) Method

Kim (2005) has developed some equations to calculate sample size for a given

power based on model fit indices CFI, RMSEA, Steiger's g , and MacDonald's fit index (Wang & Wang, 2012). Kim (2005) studied how power and minimum sample size estimates differentiated in conjunction with the fit index, the observed variables and the degrees of freedom of the model, and the covariance magnitude of variables. As Kim (2005) notes, a value of .95 for the CFI does not necessarily indicate the same misspecification as a value of .05 for the RMSEA (Kline, 2016). This happened because: 1) fit statistics tap different model fit aspects and 2) the values of fit statistics and degrees of freedom or types of model misspecification have limited correspondence (Kline, 2016 also referring Hancock & French, 2013). The resulting sample size emerging from the Kim's (2005) method is and from the Preacher and Coffman's (2006) web-based utility program for MacCallum, Browne and Sugawara's method (1996) is identical (Wang & Wang, 2012).

Finally, bootstrapping (c.f., Bollen & Stine, 1992) is another technique that also applicable to power analysis but in contrast to the rest of the methods its usefulness in determining the target N for a research is low because the generation of bootstrapped samples requires a large existing data set (Brown, 2015 referring to Jaccard & Wan, 1996).

In conclusion, the generation and inspection of power curves as functions of sample size and other assumptions is useful for planning a study. Power curves illustrate graphically the power as a function of sample size for a model (see Kline, 2016: p. 292). Statistical power can be estimated at one of two different levels in CFA/SEM. The first is the parameter level i.e. the power to detect an individual effect (Kline, 2016). An alternative level is to assess minimum required sample sizes to reach power levels equal to or greater than the desired value as Kline (2016) comments. This option is available with Monte Carlo simulation. However, the model-based approaches to power analysis have been criticized as showing low generalizability because exact estimates of population values for each parameter in the model need to be specified by the researcher (Brown, 2015).

5.6. The Bayesian Approach on Testing the Null Hypothesis

Traditional power analysis relies on testing the null hypothesis testing approach (Cohen, 1988). Nevertheless, there are alternative approaches like the Bayesian estimation approach (Wang et al., 2013). The Bayesian approach postulates that all new data is added to a sum of knowledge thus permitting the use of previous knowledge into probability determination process. In this framework hypotheses are studied by means of deductive methods using posterior probability rather than the comparison of the hypothesis examined to the null hypothesis (Barker et al., 2016).

6. What to Do When Sample Is Not Large Enough

Sometimes the sample size for a certain CFA/SEM model may not be adequate

for achieving desired power (e.g., 0.80). Nonetheless, this does not mean the researcher is left without a choice. In a SEM study with a small sample standard errors are likely to be biased and generally, the quality of goodness of fit tests may be questionable. Yet, parameter estimates are essentially unbiased if the researcher does not face non-convergence and improper solutions problems during model estimation (Chen et al., 2001). And parameter estimates are a source of useful information that can be used as guessed population inputs in a Monte Carlo simulation study on power analysis (Wang & Wang, 2012).

Additionally, Marsh and Hau (1999) offer the following guidelines for studying CFA models with a small sample size: 1) the use of indicators with good psychometric properties and with standardized coefficients $> .70$ to limit the model susceptibility to Heywood cases (Wothke, 1993). 2) The use of equality constraints on the unstandardized coefficients of indicators that belong to the same factor based on the same score limits the possibility of an inadmissible solution. This strategy is applicable to indicators having the same metric. 3) use item-parceling to analyses indicators (Kline, 2016). Also specifying models cautiously and dropping estimation of extraneous parameters is also an option (Wang et al., 2013; also quoting Floyd & Widaman, 1995).

7. Summary and Conclusions

The answer to the question “*is the sample size adequate?*” is commonly expressed by many EFA, CFA, and SEM researchers because rules of thumb were the state of the art method for years (Wang et al., 2013; Nicolaou & Masoner, 2013). Statistical power is calculated by subtracting the probability of Type II error from one. The standard limit of acceptability for statistical power is .80 i.e. 80% likelihood of rejecting a false null hypothesis (thus Type II error probability is 20% (Cohen, 1988, 1992) as Brown (2015) put it.

First, regarding EFA, literature suggested rules of thumb consisting either of minimum N s in absolute numbers like 100 - 250 (Cattell, 1978; Gorsuch, 1983), 300 (Tabachnick & Fidell, 2013) or 500 or more (Comrey & Lee, 1992) as reviewed by Dimitrov (2012). Another category of rules of thumb is ratios. In EFA the $N:p$ ratio is used, i.e. of participants (N) to variables (p) set traditionally to 5:1. However, studies suggest that strength of item loadings, uniformity of the communalities and number of items per factor (Guadagnoli & Velicer, 1988) or in two words “Strong data” (Costello & Osborne, 2005) are vital for the stability, reliability, and replicability of a factor solution (Wang et al., 2013).

Second regarding CFA and SEM the guidelines of Velicer and Fava (1998) about the size of the factor loadings and the number of variables as a function of the sample size are pertinent in CFA/SEM too. In CFA and SEM, sample size depends on a number of features like study design (e.g. cross-sectional vs. longitudinal); the number of relationships among indicators; indicator reliability, the data scaling (e.g., categorical versus continuous) and the estimator type (e.g., ML, robust ML etc.), the missing data level and pattern and model complexity

(Brown, 2015). Thus, determining sample size is approximated by power analysis (Brown, 2015; Kline, 2016; Byrne, 2012; Wang & Wang 2012). Also, minimum sample sizes are recommended to limit the non-convergence probability to have unbiased estimates or standard errors based on Monte Carlo simulations studies. Generally, CFA/SEM is a large-sample technique (Kline, 2016) but as a rule, models having robust parameter estimates and variables with high reliability may require smaller samples (Tabachnick & Fidell, 2013). Additionally, the issue whether the sample size is adequate for achieving desired power for significance tests, overall model fit, and likelihood ratio tests for specific model/research circumstances is a different aspect considered during power analysis (Hancock & French, 2013; Lee, Cai, & MacCallum, 2012). How Chi-square statistic, RMSEA, and other fit indices perform on different sample sizes levels is another parameter to consider (Hu & Bentler, 1999). Then there is sufficient power is crucial for individual parameter tests like factor loadings (Newsom, 2018). A CFA/SEM rule of thumb is the ratio of cases to free parameters, or N:q is commonly used for minimum recommendations and 10:1 to 20:1 is a commonly suggested ratio (Schumacker & Lomax, 2015; Kline, 2016; Jackson, 2003). Anyhow, even suggestions based on simulation studies are only rough approximations, not equally applicable to all SEM studies. Simulation studies have the potential to study only a fraction of SEM research conditions at a time thus they are not easily generalized (Brown, 2015; Newsom, 2018).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

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. <https://doi.org/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. <https://doi.org/10.1037/0033-2909.103.3.411>
- Bandalos, D. L. (2014). Relative Performance of Categorical Diagonally Weighted Least Squares and Robust Maximum Likelihood Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 102-116. <https://doi.org/10.1080/10705511.2014.859510>
- Bandalos, D. L., & Leite, W. (2013). Use of Monte Carlo Studies in Structural Equation Modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2nd ed., pp. 625-666). Charlotte, NC: IAP.
- Barker, C., Pistrang, N., & Elliott, R. (2015). *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners* (3rd ed.). Oxford, UK: John Wiley & Sons, Ltd.
- Bentler, P. M., & Chou, C. P. (1987). Practical Issues in Structural Modeling. *Sociological Methods & Research*, *16*, 78-117. <https://doi.org/10.1177/0049124187016001004>

- Bentler, P. M., & Yuan, K. H. (1999). Structural Equation Modeling with Small Samples: Test Statistics. *Multivariate Behavioral Research, 34*, 181-197.
<https://doi.org/10.1207/S15327906Mb340203>
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Jon Wiley & Sons. <https://doi.org/10.1002/9781118619179>
- Bollen, K. A., & Liang, J. (1988). Some Properties of Hoelter's CN. *Sociological Methods & Research, 16*, 492-503. <https://doi.org/10.1177/0049124188016004003>
- Bollen, K. A., & Stine, R. A. (1992). Bootstrapping Goodness-of-Fit Measures in Structural Equation Models. *Sociological Methods & Research, 21*, 205-229.
<https://doi.org/10.1177/0049124192021002004>
- Boomsma, A. (1985). Nonconvergence, Improper Solutions, and Starting Values in LISREL Maximum Likelihood Estimation. *Psychometrika, 50*, 229-242.
<https://doi.org/10.1007/BF02294248>
- Boomsma, A., & Hoogland, J. J. (2001). The Robustness of LISREL Modeling Revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural Equation Models: Present and Future. A Festschrift in Honor of Karl Jöreskog* (pp. 139-168). Lincolnwood, IL: Scientific Software International.
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York: The Guilford Press.
- Byrne, B. M. (2012). *Structural Equation Modeling with Mplus. Basic Concepts, Applications, and Programming* (2nd ed.). New York: Routledge.
- Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum. <https://doi.org/10.1007/978-1-4684-2262-7>
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper Solutions in Structural Equation Models: Causes, Consequences, and Strategies. *Sociological Methods and Research, 29*, 468-508. <https://doi.org/10.1177/0049124101029004003>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling, 9*, 233-255.
https://doi.org/10.1207/S15328007SEM0902_5
- Chumney, F. L. (2013). *Structural Equation Models with Small Samples: A Comparative Study of Four Approaches* (p. 189). College of Education and Human Sciences.
<http://digitalcommons.unl.edu/cehsdiss/189>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*.
- Cohen, J. (1990). Things I Have Learned (so far). *American Psychologist, 45*, 1304-1312.
<https://doi.org/10.1037/0003-066X.45.12.1304>
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112*, 155-159.
<https://doi.org/10.1037/0033-2909.112.1.155>
- Comrey, A. L. (1988). Factor-Analytic Methods of Scale Development in Personality and Clinical Psychology. *Journal of Consulting and Clinical Psychology, 56*, 754-761.
<https://doi.org/10.1037/0022-006X.56.5.754>
- Comrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Comrey, A. L., & Lee, H. B. (1992). *Interpretation and Application of Factor Analytic Results*.
- Comrey, A. L., Backer, T. E., & Glaser, E. M. (1973). *A Sourcebook for Mental Health Measures*.
- Coolican, H. (2014). *Research Methods and Statistics in Psychology* (6th ed.). New York, NY: Psychology Press.

- Costello, A. B., & Osborne, J. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis. *Practical Assessment Research & Evaluation, 10*, 1-9.
- 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. https://doi.org/10.1207/S15327906MBR3701_01
- DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). Thousand Oaks, CA: Sage.
- Dimitrov, D. M. (2012). *Statistical Methods for Validation of Assessment Scale Data in Counseling and Related Fields*. Alexandria, VA: American Counseling Association.
- 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: A Multidisciplinary Journal, 2*, 119-143. <https://doi.org/10.1080/10705519509540000>
- Du, H., Zhang, Z., & Yuan, K. (2017). Power Analysis for t-Test with Non-normal Data and Unequal Variances. In L. A. van der Ark et al. (Eds.), *Quantitative Psychology, Springer Proceedings in Mathematics & Statistics* (Volume 196, pp.373-380). Switzerland: Springer International. https://doi.org/10.1007/978-3-319-56294-0_32
- Everitt, B. S. (1975). Multivariate Analysis: The Need for Data, and Other Problems. *The British Journal of Psychiatry, 126*, 237-240. <https://doi.org/10.1192/bjp.126.3.237>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods, 4*, 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Fan, X., & Sivo, S. A. (2007). Sensitivity of Fit Indices to Model Misspecification and Model Types. *Multivariate Behavioral Research, 42*, 509-529. <https://doi.org/10.1080/00273170701382864>
- Finch, H. W., Immekus, J. C., & French, B. F. (2016). *Applied Psychometrics Using SPSS and AMOS*. Charlotte, NC: Information Age Publishing Inc.
- Floyd, F. J., & Widaman, K. F. (1995). Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment, 7*, 286-299. <https://doi.org/10.1037/1040-3590.7.3.286>
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling, 16*, 625-641. <https://doi.org/10.1080/10705510903203573>
- Gatignon, H. (2010). Confirmatory Factor Analysis. In *Statistical Analysis of Management Data* (pp. 59-122). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-1270-1_4
- Gorsuch, R. (1983). *Factor Analysis*. Hillsdale, NJ: L. Erlbaum Associates.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of Sample Size to the Stability of Component Patterns. *Psychological Bulletin, 103*, 265-275. <https://doi.org/10.1037/0033-2909.103.2.265>
- Hancock, G. R., & French, B. F. (2013). Power Analysis in Structural Equation Modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2nd ed., pp. 117-159). Charlotte, NC: IAP.
- Hau, K. T., & Marsh, H. W. (2004). The Use of Item Parcels in Structural Equation Modeling: Non-Normal Data and Small Sample Sizes. *British Journal of Mathematical and Statistical Psychology, 57*, 327-351. <https://doi.org/10.1111/j.2044-8317.2004.tb00142.x>

- Hoe, S. L. (2008). Issues and Procedures in Adopting Structural Equation Modeling Technique. *Journal of Applied Quantitative Methods*, 3, 76-83.
- Hoelter, J. W. (1983). The Analysis of Covariance Structures: Goodness-of-Fit Indices. *Sociological Methods & Research*, 11, 325-344.
<https://doi.org/10.1177/0049124183011003003>
- Hoogland, J. J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research*, 26, 329-367. <https://doi.org/10.1177/0049124198026003003>
- Hoyle, R. H. (1999). *Statistical Strategies for Small Sample Research*. New York, NY: Sage.
- Hoyle, R. H., & Kenny, D. A. (1999). Statistical Power and Tests of Mediation. In R. H. Hoyle (ed.), *Statistical Strategies for Small Sample Research* (pp. 195-222). New York, NY: SAGE Publications.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55. <https://doi.org/10.1080/10705519909540118>
- Jaccard, J., & Wan, C. K. (1996). *LISREL Approaches to Interaction Effects in Multiple Regression* (No. 114). New York, NY: SAGE Publications.
<https://doi.org/10.4135/9781412984782>
- Jackson, D. L. (2001). Sample Size and Number of Parameter Estimates in Maximum Likelihood Confirmatory Factor Analysis: A Monte Carlo Investigation. *Structural Equation Modeling*, 8, 205-223. https://doi.org/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*, 10, 128-141.
https://doi.org/10.1207/S15328007SEM1001_6
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International.
- Kelloway, E. K. (2015). *Using Mplus for Structural Equation Modeling*. Thousand Oaks, CA: Sage.
- Kim, K. H. (2005). The Relation among Fit Indexes, Power, and Sample Size in Structural Equation Modeling. *Structural Equation Modeling*, 12, 368-390.
https://doi.org/10.1207/s15328007sem1203_2
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.).
- Lee, T., Cai, L., & MacCallum, R. (2012). Power Analysis for Tests of Structural Equation Models. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 181-194). New York: Guilford Press.
- Loehlin, J. C. (2004). *Latent Variable Models* (4th ed.). Mahwah, NJ: Erlbaum.
- Loehlin, J. C., & Beaujean, A. A. (2017). *Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis*. New York, NY: Taylor & Francis.
- Lomax, R. G., & Hahs-Vaughn, D. L. (2013). *An Introduction to Statistical Concepts*. Abingdon-on-Thames: Routledge.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, 51, 201-226.
<https://doi.org/10.1146/annurev.psych.51.1.201>
- MacCallum, R. C., & Hong, S. (1997). Power Analysis in Covariance Structure Modeling Using GFI and AGFI. *Multivariate Behavioral Research*, 32, 193-210.
https://doi.org/10.1207/s15327906mbr3202_5

- 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. <https://doi.org/10.1037/1082-989X.1.2.130>
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample Size in Factor Analysis. *Psychological Methods, 4*, 84-99. <https://doi.org/10.1037/1082-989X.4.1.84>
- MacCallum, R., Lee, T., & Browne, M. W. (2010). The Issue of Isopower in Power Analysis for Tests of Structural Equation Models. *Structural Equation Modeling, 17*, 23-41. <https://doi.org/10.1080/10705510903438906>
- Marsh, H. W., & Hau, K. T. (1999). Confirmatory Factor Analysis: Strategies for Small Sample Sizes. *Statistical Strategies for Small Sample Research, 1*, 251-284.
- 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. https://doi.org/10.1207/s15327906mbr3302_1
- Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction. *Psychological Methods, 9*, 275-300. <https://doi.org/10.1037/1082-989X.9.3.275>
- McDonald, R. P., & Marsh, H. W. (1990). Choosing a Multivariate Model: Noncentrality and Goodness of Fit. *Psychological Bulletin, 107*, 247-255. <https://doi.org/10.1037/0033-2909.107.2.247>
- McQuitty, S. (2004). Statistical Power and Structural Equation Models in Business Research. *Journal of Business Research, 57*, 175-183. [https://doi.org/10.1016/S0148-2963\(01\)00301-0](https://doi.org/10.1016/S0148-2963(01)00301-0)
- Mulaik, S. A. (1990). Blurring the Distinctions between Component Analysis and Common Factor Analysis. *Multivariate Behavioral Research, 25*, 53-59. https://doi.org/10.1207/s15327906mbr2501_6
- 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, 9*, 599-620. https://doi.org/10.1207/S15328007SEM0904_8
- Nevitt, J., & Hancock, G. R. (2001). Performance of Bootstrapping Approaches to Model Test Statistics and Parameter Standard Error Estimation in Structural Equation Modeling. *Structural Equation Modeling, 8*, 353-377. https://doi.org/10.1207/S15328007SEM0803_2
- Nevitt, J., & Hancock, G. R. (2004). Evaluating Small Sample Approaches for Model Test Statistics in Structural Equation Modeling. *Multivariate Behavioral Research, 39*, 439-478. https://doi.org/10.1207/S15327906MBR3903_3
- Newsom, J. T. (2018). *Minimum Sample Size Recommendations (Psy 523/623 Structural Equation Modeling, Spring 2018)*. Manuscript Retrieved from upa.pdx.edu/IOA/newsom/semrefs.htm.
- Nicolaou, A. I., & Masoner, M. M. (2013). Sample Size Requirements in Structural Equation Models under Standard Conditions. *International Journal of Accounting Information Systems, 14*, 256-274. <https://doi.org/10.1016/j.accinf.2013.11.001>
- Nunnally, J. C., & Bernstein, I. H. (1967). *Psychometric Theory* (Vol. 226). New York, NY: McGraw-Hill.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (McGraw-Hill Series in Psychology, Vol. 3). New York, NY: McGraw-Hill.
- Preacher, K. J., & Coffman, D. L. (2006). *Computing Power and Minimum Sample Size for RMSEA*. <http://quantpsy.org>

- Raykov, T. (2012). Scale Construction and Development Using Structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 472-492). New York, NY: Guilford Press.
- Saris, W. E., & Satorra, A. (1993). Power Evaluations in Structural Equation Models. In K. A. Bollen, & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 181-204). Newbury Park, CA: Sage.
- Saris, W. E., & Stronkhorst, L. H. (1984). *Causal Modelling in Nonexperimental Research: An Introduction to the LISREL Approach* (Vol. 3). Sociometric Research Foundation.
- Saris, W. E., Satorra, A., & Van der Veld, W. M. (2009). Testing Structural Equation Models or Detection of Misspecifications? *Structural Equation Modeling*, *16*, 561-582.
- Satorra, A., & Saris, W. E. (1985). Power of the Likelihood Ratio Test in Covariance Structure Analysis. *Psychometrika*, *50*, 83-90. <https://doi.org/10.1007/BF02294150>
- Schumacker, R. E., & Lomax, R. G. (2015). *A Beginner's Guide to Structural Equation Modeling* (4th ed.). New York, NY: Routledge.
- Schumacker, R. E., & Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling* (4th ed.). New York: Routledge.
- Silvia, E. S. M., & MacCallum, R. C. (1988). Some Factors Affecting the Success of Specification Searches in Covariance Structure Modeling. *Multivariate Behavioral Research*, *23*, 297-326. https://doi.org/10.1207/s15327906mbr2303_2
- Singh, K., Junnarkar, M., & Kaur, J. (2016). *Measures of Positive Psychology, Development and Validation*. Berlin: Springer
- Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics*. Boston, MA: Pearson Education Inc.
- 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. <https://doi.org/10.2307/1130296>
- Thomas, L. (1997). Retrospective Power Analysis. *Conservation Biology*, *11*, 276-280. <https://doi.org/10.1046/j.1523-1739.1997.96102.x>
- Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. Washington DC: American Psychological Association.
- Tinsley, H. E., & Tinsley, D. J. (1987). Uses of Factor Analysis in Counseling Psychology Research. *Journal of Counseling Psychology*, *34*, 414-424. <https://doi.org/10.1037/0022-0167.34.4.414>
- Velicer, W. F., & Fava, J. L. (1998). Affects of Variable and Subject Sampling on Factor Pattern Recovery. *Psychological Methods*, *3*, 231-251. <https://doi.org/10.1037/1082-989X.3.2.231>
- Wang, J., & Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus*. Hoboken, NJ: Wiley, Higher Education Press. <https://doi.org/10.1002/9781118356258>
- Wang, L. L., Watts, A. S., Anderson, R. A., & Little, T. D. (2013). Common Fallacies in Quantitative Research Methodology. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (pp. 718-758). New York: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199934898.013.0031>
- Wheaton, B. (1987). Assessment of Fit in Overidentified Models with Latent Variables. *Sociological Methods & Research*, *16*, 118-154. <https://doi.org/10.1177/0049124187016001005>
- Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. F. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology*, *8*, 84-136. <https://doi.org/10.2307/270754>

- Widaman, K. F. (1993). Common Factor Analysis versus Principal Components Analysis: Differential Bias in Representing Model Parameters. *Multivariate Behavioral Research, 28*, 263-311. https://doi.org/10.1207/s15327906mbr2803_1
- Wilcox, R. R. (2008). Sample Size and Statistical Power. In A. M. Nezu, & C. M. Nezu (Eds.), *Evidence-Based Outcome Research: A Practical Guide to Conducting Randomized Controlled Trials for Psychosocial Interventions* (pp. 123-134). New York, NY: Oxford University Press.
- Williams, B., Onsmann, A., & Brown, T. (2010). Exploratory Factor Analysis: A Five-Step Guide for Novices. *Australasian Journal of Paramedicine, 8*, 1-13.
- 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. <https://doi.org/10.1177/0013164413495237>
- Wothke, W. (1993). Nonpositive Definite Matrices in Structural Modeling. *Sage Focus Editions, 154*, 256-256.
- Yu, C.-Y., & Muthén, B. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes*. Doctoral Dissertation. <http://www.statmodel.com/download/Yudissertation.pdf>
- Yuan, K.-H., & Bentler, P. M. (2000). Three Likelihood-Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data. *Sociological Methodology, 30*, 165-200. <https://doi.org/10.1111/0081-1750.00078>