

# Sample Size Requirements for Simple and Complex Mediation Models

Educational and Psychological  
Measurement

2022, Vol. 82(1) 76–106

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00131644211003261

journals.sagepub.com/home/epm



Mikyung Sim<sup>1</sup> , Su-Young Kim<sup>1</sup>, and Youngsuk Suh<sup>2</sup> 

## Abstract

Mediation models have been widely used in many disciplines to better understand the underlying processes between independent and dependent variables. Despite their popularity and importance, the appropriate sample sizes for estimating those models are not well known. Although several approaches (such as Monte Carlo methods) exist, applied researchers tend to use insufficient sample sizes to estimate their models of interest, which might result in unstable and inaccurate estimation of the model parameters including mediation effects. In the present study, sample size requirements were investigated for four frequently used mediation models: one simple mediation model and three complex mediation models. For each model, path and structural equation modeling approaches were examined, and partial and complete mediation conditions were considered. Both the percentile bootstrap method and the multivariate delta method were compared for testing mediation effects. A series of Monte Carlo simulations was conducted under various simulation conditions, including those concerning the level of effect sizes, the number of indicators, the magnitude of factor loadings, and the proportion of missing data. The results not only present practical and general guidelines for substantive researchers to determine minimum required sample sizes but also improve understanding of which factors are related to sample size requirements in mediation models.

## Keywords

mediation analysis, mediation model, sample size, indirect effect, bootstrap method

---

<sup>1</sup>Ewha Womans University, Seoul, Republic of Korea

<sup>2</sup>Korean Education and Psychology Institute, Seoul, Republic of Korea

## Corresponding Author:

Youngsuk Suh, Korean Education and Psychology Institute, Mapo-gu Mapo-daero 173, Seoul 04130, Republic of Korea.

Email: yssuh327@gmail.com

Mediation analysis has been widely used for decades to better understand the relationship between independent and dependent variables and recently has become one of the most popular statistical models in both methodological studies (e.g., Fritz & MacKinnon, 2007; Lachowicz et al., 2018; Liu & Wang, 2019; B. Muthén & Asparouhov, 2015; Thoemmes et al., 2010) and substantive research (e.g., Duckworth et al., 2016; Lockman & Servaty-Seib, 2018; Malone et al., 2016). In a mediation model, the effect of an independent variable on a dependent variable is explained through a third variable called a mediator (Fritz & MacKinnon, 2007). In many different disciplines, mediation models have been utilized to answer various research questions. For example, in psychology, the associations between midlife Eriksonian development and both late-life global cognition and executive functioning were partially mediated by late-life depression (Malone et al., 2016). In business management, the effect of supplier-facing purchasing and supply management practices on operational performance was fully mediated by internal purchasing and supply management practices (Foerstl et al., 2016). In health science, the effect of past mental health on present physical health was intervened through lifestyle choices and social interactions (Ohrnberger et al., 2017).

Although mediation models provide useful tools for describing the relationship between a stimulus and a response (MacKinnon, 2008), a common issue encountered by substantive researchers has been to determine appropriate sample sizes for estimating mediation effects properly. It has been well known that the standard error of a test statistic is affected by sample size, which in turn, affects the power and/or Type I error rate. This implies that despite statistical nonsignificance due to smaller sample sizes than needed (in terms of appropriate power, accurate estimation, etc.), a non-ignorable effect size may be present. On the other hand, even when statistical significance is obtained using a test, the effect may be negligible or without any practical importance, especially when the sample size is much larger than needed. Too large a sample size could thus squander the cost, time, and effort of conducting research.

Determining proper sample sizes is one of the unresolved, important issues in the utilization of mediation models. Although L. K. Muthén and Muthén (2002) showed how to perform Monte Carlo simulations to determine required sample size given a statistical model, Monte Carlo simulation is still not an accessible method for most applied researchers. Therefore, there have been several studies investigating sample size requirements, each with its strengths and limitations. For example, Fritz and MacKinnon (2007) conducted Monte Carlo simulations regarding the performance of six mediation testing methods (Sobel's first-order test [Sobel, 1982], the percentile bootstrap test [Shrout & Bolger, 2002], etc.), which were commonly employed in the literature and provided the sample sizes necessary for achieving .8 power in a single-mediator path model with several effect size combinations of path coefficients. Similarly, Liu and Wang (2019) presented a method of planning sample sizes for a simple<sup>1</sup> mediation path model. They conducted a simulation study to examine the impact of uncertainty in effect size estimates on the power of the joint significance test (Cohen & Cohen, 1983). *R* functions and a web application were developed for

implementing the proposed method. The researchers chose the levels of effect sizes following Fritz and MacKinnon (2007) so that the results could be easily compared. However, as previously mentioned, both studies were limited to only a simple mediation path model. Wolf et al. (2013) evaluated sample size requirements for confirmatory factor analysis (CFA) models with one, two, and three factors, and simple mediation models with and without latent variables via Monte Carlo simulations. The authors first examined the impact of key model properties, such as the number of indicators and the magnitude of factor loadings on CFA models, and then they held those components constant (i.e., three indicators and 0.65 factor loading) to examine the impact of changes in effect sizes for structural path coefficients in mediation models. They also considered the effect of missing data on sample size requirements. Although their results provided some guidelines for sample size requirements on mediation models, only simple mediation models were considered with limited conditions. For more complex types of mediation, Schoemann et al. (2017) proposed a new method along with a convenient tool (a web application using *R*) for determining power and sample sizes in simple- and multiple-mediator (i.e., two-parallel or two-serial mediators) models. Although their application provided researchers with an easy-to-use tool in designing mediation studies, the method could be applied only to certain types of path model, not to structural equation models (SEMs).

Other than the sample size requirement studies listed above, there are a few other studies that have focused on power analysis in designing mediation models. For example, Thoemmes et al. (2010) described a general framework for estimating power based on the L. K. Muthén and Muthén's (2002) Monte Carlo approach for simple and complex mediation models that include multiple mediators, three-path mediation, mediation with latent variables, moderated mediation, and mediation in longitudinal designs. They provided tabled values of required sample sizes for some models but under limited simulation conditions. Also, only mediating variables were considered as latent variables, which seems less likely to occur in practice. Similarly, Zhang (2014) proposed a power analysis to detect mediation effects based on the percentile bootstrap method via Monte Carlo simulations. The method can handle nonnormal data with excessive skewness and kurtosis, and some examples that utilized complex mediation models (e.g., a multigroup model and a longitudinal model) were illustrated. An *R* package was also developed for its application.

As discussed, several studies have investigated sample sizes in mediation analysis (e.g., Fritz & MacKinnon, 2007; Liu & Wang, 2019; Schoemann et al., 2017). However, the suggestions from the existing body of work are limited to a narrow range of models, such as simple mediation models or path models. Few studies included SEMs, but even these did not consider the various analytic factors that are associated with mediation models (e.g., the number of indicators, the magnitude of factor loadings, etc.). In addition, the majority of previous studies focused on statistical power as a criterion in determining sample sizes. According to the sample size study by L. K. Muthén and Muthén (2002), there are several additional criteria (e.g., parameter bias and 95% coverage) that can be considered other than just power.

Given the limitations of the previous methodological studies on sample size requirements, we first conducted a systematic search of the literature to investigate the actual sample sizes used in substantive research on mediation analyses. Our review covered articles published between 2016 and 2018 in some psychology journals including *Journal of Counseling Psychology*, *Developmental Psychology*, *Journal of Applied Psychology*, and *Journal of Educational Psychology*. A total of 2,562 articles were published in the journals. We searched articles with the key words, “mediation,” “mediating,” “mediated,” or “indirect effect” in the abstract, using PsyINFO and identified 355 articles. We then eliminated articles that did not include the subject of mediational tests and those that concerned multilevel mediation and/or longitudinal mediation. Finally, 201 articles were examined, focusing on the mediation models and sample sizes used in them. The four most frequently used models identified in the articles are displayed in Figure 1. To save space, path diagrams for structural equation mediation models only with three indicators are displayed. In this study, Model 1 is called the simple mediation model, and the other three models are called complex mediation models. Specifically, Models 2 and 4 are multiple mediation models, whereas Model 3 is a multiple-step multiple mediation model (Hayes, 2009). In other words, Models 2 and 4 have two simple mediation effects in each model, and Model 3 includes a multiple-step mediation effect (i.e.,  $a_1db_2$ ) besides two simple mediation effects ( $a_1b_1$  and  $a_2b_2$ ). Detailed explanations about the mediation models are provided in the next section.

Table 1 shows the frequency (percentage) of mediation models (path models, SEMs, and the total) and sample sizes used for testing mediation effects in them. It should be noted that some articles employed more than one model, and thus, the total in Table 1 is 206 not 201. The most frequently used model was Model 1 (33.98%) of the total, and Model 3 was next (9.22%). Models 1 to 4 comprised 58.25% of the total, and the remainder of the studies used other types of models (e.g., complex models with multiple-dependent variables and/or multiple mediators). In addition, 57.28% of the total were SEMs, while the rest were path models. The sample sizes used for the models were summarized as minimum, maximum, mean, and median in Table 1. The smallest sample size was 73, and the largest sample size was 13,645. The median sample sizes for path models ranged from 171 to 308, while those for SEMs ranged from 284 to 368. Although not shown in Table 1, 74.31% of the total were involved in partial mediation effects, and the rest were associated with complete mediation effects.

Considering the popularity of mediation analyses in substantive research, it is worthwhile to verify whether the sample sizes used in practice are appropriate and to investigate the minimum sample size requirements for frequently used mediation models. The main purpose of the present study is to provide substantive researchers with practical and general guidelines for determining sample sizes in the four commonly used mediation models, as shown in Figure 1: one simple mediation model and three complex mediation models. We compared SEMs with path models to investigate the effect of including latent variables on sample size requirements. For each modeling approach, both partial and complete mediation conditions were

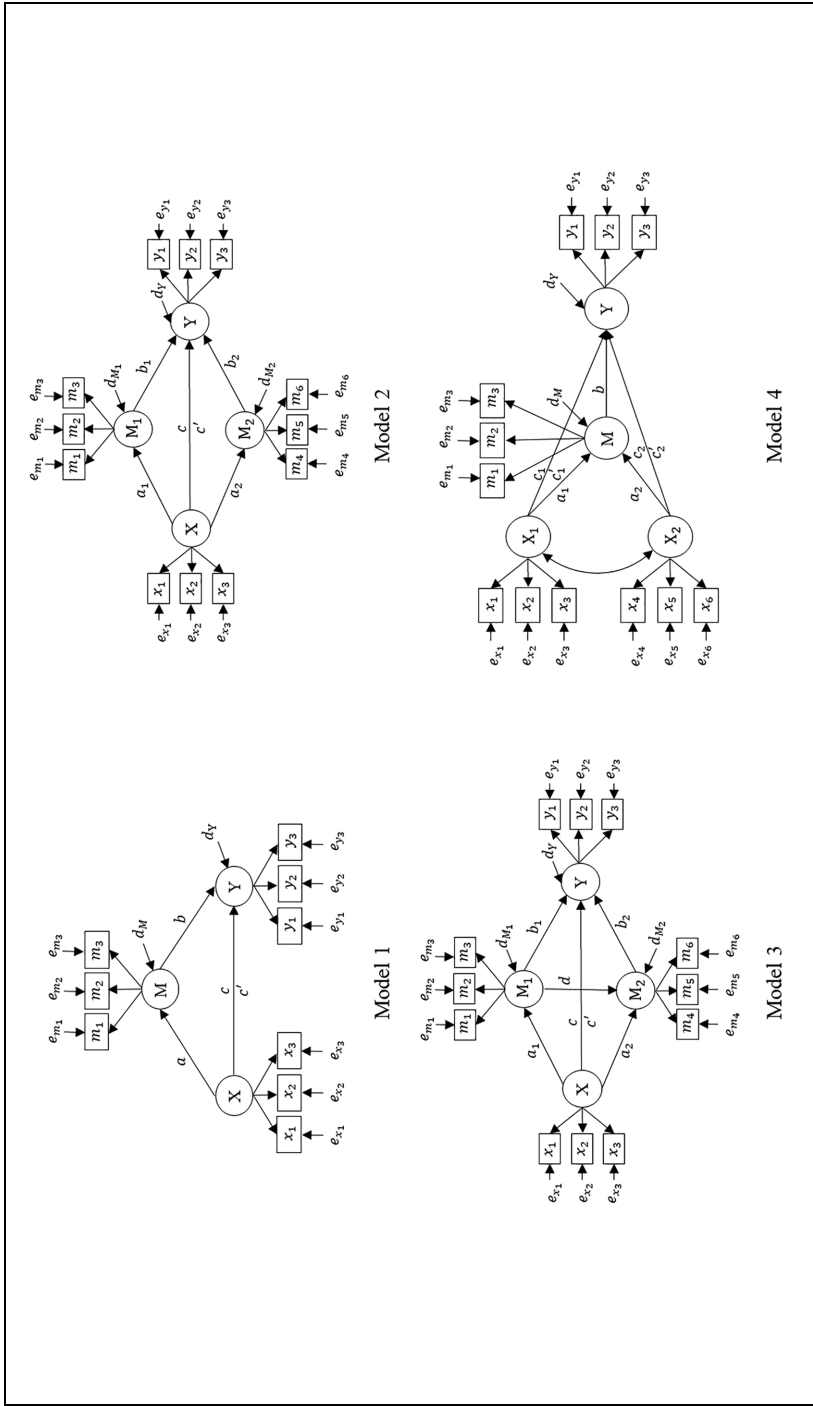


Figure 1. Mediation models.

**Table 1.** Review of the Literature on Mediation Models and Sample Sizes in Research Applications.

Mediation model	Frequency	%	Sample size				
			Minimum	Maximum	M	Mdn	
Path model	Model 1	43	48.86	86	2,451	340	242
	Model 2	7	7.95	128	8,879	1,442	171
	Model 3	3	3.41	270	347	297	274
	Model 4	5	5.68	141	1,851	575	299
	Other <sup>a</sup>	30	34.09	93	1,042	347	308
	Total	88	100.00	86	8,879	447	266
SEM	Model 1	27	22.88	73	645	334	316
	Model 2	8	6.78	208	2,088	585	356
	Model 3	16	13.56	221	1,000	428	365
	Model 4	11	9.32	148	545	333	284
	Other	56	47.46	111	13,645	700	368
	Total	118	100.00	73	13,645	538	352
Total	Model 1	70	33.98	73	2,451	335	293
	Model 2	15	7.28	128	8,879	950	317
	Model 3	19	9.22	221	1,000	409	352
	Model 4	16	7.77	141	1,851	409	292
	Other	86	41.75	93	13,645	582	335
	Total	206	100.00	73	13,645	500	311

Note. SEM = structural equation model. M = mean. Mdn = median.

<sup>a</sup>Other types of models include complex models with multiple-dependent variables and/or multiple mediators.

considered to reflect the recent practice found in the literature search. Also, the percentile bootstrap method (Shrout & Bolger, 2002) and the multivariate delta method (MacKinnon, 2008; Sobel, 1982) were utilized to detect mediation effects, and their performances were compared. A series of Monte Carlo simulations was conducted under various simulation conditions. In particular, the number of indicators, the magnitude of factor loadings, and the proportion of missing data for the SEM approach were included as simulation factors. To our knowledge, these have not been extensively studied in the literature on mediation.

The remaining sections of the article proceed as follows. In the next section, a description of the mediation models used in this study is provided, along with mediation effects. In the following section, a Monte Carlo simulation study is presented in terms of simulation design and data analysis. Next, simulation results are separately provided for path models and SEMs. In the final section, a brief summary of findings is presented, and the implications of the results are discussed.

### Mediation Models

Figure 1 displays the four mediation models with indicators (i.e., SEMs for mediation) used in this study. Model 1 is the simple mediation model, which involves three

latent variables (i.e.,  $X$ ,  $Y$ , and  $M$ ) and examines if the effect of an independent variable ( $X$ ) on a dependent variable ( $Y$ ) is intervened through a mediator ( $M$ ). In Model 1,  $a$  is the effect of  $X$  on  $M$ ,  $b$  is the effect of  $M$  on  $Y$ , and  $c'$  is the direct effect of  $X$  on  $Y$  adjusted for  $M$ . The product of  $a$  and  $b$  ( $ab$ ) is referred to as a mediation effect or an indirect effect.  $d_M$  and  $d_Y$  are residual or disturbance terms in the structural part of the model, which are usually assumed to be normally distributed with a mean of zero.  $m_1$  to  $m_3$ ,  $x_1$  to  $x_3$ , and  $y_1$  to  $y_3$  are indicators of latent constructs, and the corresponding  $e$  terms are the errors in the measurement part of the model. If only  $X$  and  $Y$  were present in the model without  $M$ ,  $c$  would indicate the total effect of  $X$  on  $Y$ . The total effect ( $c$ ) is equal to the sum of the direct effect ( $c'$ ) and indirect effect ( $ab$ ). Complete mediation occurs when  $X$  no longer affects  $Y$  after  $M$  is held constant, making  $c'$  statistically zero, while partial mediation is present when  $c'$  is less than  $c$ , but is still different from zero (Baron & Kenny, 1986).<sup>2</sup> Models 2 and 3 are complex mediation models with two mediators ( $M_1$  and  $M_2$ ). Model 2 shows two simple mediation effects,  $a_1b_1$  and  $a_2b_2$ , while Model 3 presents three mediation effects, that is, two single-step mediation effects ( $a_1b_1$  and  $a_2b_2$ ) and a multiple-step mediation effect ( $a_1db_2$ ). Model 4 is a mediation model with two independent variables and shows two single-step mediation effects,  $a_1b$  and  $a_2b$ . For each model, a corresponding path model (without latent variables and indicators) is also considered in this study.

In the present investigation, all variables (i.e., indicators and latent variables for SEMs and measured variables for path models) are assumed to be continuous and standardized (i.e., mean of zero and variance of one) as in other simulation studies (e.g., Falk & Biesanz, 2015; Gagné & Hancock, 2006; Jackson et al., 2013). Due to the use of the standardized variables, the path coefficients are also standardized, as are the indirect effects. The reasons for using the standardized indirect effects are as follows. They have desirable statistical properties, such as invariant to proper linear transformations, independent of sample sizes (Lachowicz et al., 2018), and scale-free. Besides, standardized indirect effects are frequently used in substantive mediation studies (Liu & Wang, 2019).

## Method

### *Study Design and Data Generation*

A Monte Carlo simulation study was conducted to examine sample size requirements for the introduced simple and complex mediation models. Under each model, path coefficients were simulated to indicate three different levels of effect sizes. Table 2 displays the selected path coefficients and variance parameters under the three indirect effect size conditions for each model. For the simple mediation model (Model 1 in Figure 1), three magnitudes, 0.14, 0.36, and 0.51, were considered for  $a$  and  $b$ , corresponding to Cohen's (1988)  $R^2$  criteria for small (2% of the variance in the dependent variable that is explained), medium (13% of the variance), and large (26% of the variance) effect sizes, respectively. These values were selected to resemble those used

**Table 2.** Path Coefficients and Residual Variance Parameters Under Three Indirect Effect Size Conditions.

Model	Parameter	Partial mediation			Complete mediation		
		Small	Medium	Large	Small	Medium	Large
Model 1	$a$	0.14	0.36	0.51	0.14	0.36	0.51
	$b$	0.14	0.36	0.51	0.14	0.36	0.51
	$c'$	0.14	0.14	0.14	0.00	0.00	0.00
	$Var(d_M)$	0.98	0.87	0.74	0.98	0.87	0.74
	$Var(d_Y)$	0.95	0.81	0.65	0.98	0.87	0.74
Model 2	$a_1, a_2$	0.14	0.36	0.51	0.14	0.36	0.51
	$b_1, b_2$	0.09	0.25	0.41	0.10	0.27	0.42
	$c'$	0.04	0.14	0.14	0.00	0.00	0.00
	$Var(d_{M_1})$	0.98	0.87	0.74	0.98	0.87	0.74
	$Var(d_{M_2})$	0.98	0.87	0.74	0.98	0.87	0.74
Model 3	$Var(d_Y)$	0.98	0.80	0.53	0.98	0.85	0.65
	$a_1$	0.14	0.36	0.51	0.14	0.36	0.51
	$a_2$	0.10	0.27	0.42	0.10	0.27	0.42
	$d$	0.10	0.27	0.42	0.10	0.27	0.42
	$b_1$	0.09	0.25	0.41	0.10	0.27	0.42
	$b_2$	0.09	0.25	0.41	0.10	0.27	0.42
	$c'$	0.04	0.14	0.14	0.00	0.00	0.00
	$Var(d_{M_1})$	0.98	0.87	0.74	0.98	0.87	0.74
Model 4	$Var(d_{M_2})$	0.98	0.80	0.47	0.98	0.80	0.47
	$Var(d_Y)$	0.98	0.80	0.54	0.98	0.85	0.65
	$a_1, a_2$	0.10	0.27	0.42	0.10	0.27	0.42
	$b$	0.13	0.33	0.56	0.14	0.36	0.51
	$c'_1, c'_2$	0.04	0.14	0.14	0.00	0.00	0.00
	$Var(d_M)$	0.97	0.81	0.55	0.97	0.81	0.55
	$Var(d_Y)$	0.98	0.79	0.51	0.98	0.87	0.74

in previous studies (e.g., Fritz & MacKinnon, 2007; Liu & Wang, 2019; Thoemmes et al., 2010). For  $c'$ , only two magnitudes, 0 and 0.14, were used to manipulate the complete mediation condition<sup>3</sup> and the partial mediation (a small direct effect size) condition, respectively (Liu & Wang, 2019). Once the path coefficients were chosen, the error variances of  $d_M$  and  $d_Y$  could be calculated. Calculation examples are shown in Appendix A.

For the three complex mediation models (Models 2 to 4), Cohen's (1988)  $R^2$  were used again to simulate the three effect size conditions. In Models 2 and 3, two mediators were included as endogenous variables, and the residual covariance between  $d_{M_1}$  and  $d_{M_2}$  was fixed to 0 for simplicity. For Model 4, 0.3 was used as the parameter value for the covariance between  $X_1$  and  $X_2$  regardless of the effect sizes, following the value used for the correlation between factors considered in previous studies (Gagné & Hancock, 2006; Marsh et al., 1998; Wolf et al., 2013). In addition, model-specific path coefficients were selected as follows. In Model 2, the amount of



explained variance in  $M_1$  and  $M_2$  were determined by  $X$  only, and thus the values for the corresponding path coefficients,  $a_1$  and  $a_2$  were selected for each of the three effect size conditions as in the simple mediation model. The amount of explained variance in  $Y$  was determined by  $X$ ,  $M_1$ , and  $M_2$ , which were related to three path coefficients,  $c'$ ,  $b_1$ , and  $b_2$ , respectively. For simplicity,  $M_1$  and  $M_2$  were assumed to contribute equally, and  $X$  was presumed to have a small effect on  $Y$  (i.e., the explained variance of 2% and thus,  $c' = 0.14$  as in Model 1) under the medium and large effect size conditions. For the small effect size (2%) condition, assigning the whole 2% to  $X$  made the other two path coefficients equal to zero ( $b_1 = b_2 = 0$ ). Therefore, only 0.2% was arbitrarily assigned to  $X$ , creating  $c' = 0.04$ . The complete mediation condition ( $c' = 0$ ) was also simulated. In Model 3, the amount of explained variance in  $M_2$  was determined by  $X$  and  $M_1$ , which were assumed to contribute equally (i.e.,  $a_2 = d$ ).  $c'$ ,  $b_1$ , and  $b_2$  were set following the way implemented in Model 2. Finally,  $X_1$  and  $X_2$  were involved in Model 4, and they were assumed to have the same effect on  $M$  (i.e.,  $a_1 = a_2$ ) and on  $Y$  (i.e.,  $c'_1 = c'_2$ ).  $c'_1$  and  $c'_2$  were determined like  $c'$  in Model 2. In each of the complex mediation models, the error variances were also calculated in accordance with the selected values for the path coefficients.

As mentioned earlier, two modeling approaches (i.e., SEMs and path models) were considered in this study. In both SEMs and path models, the same effect size conditions were applied. Unlike path models, the latent constructs in SEMs are measured by indicators, assuming the existence of measurement errors. Hence, two more simulation factors relating to indicators were included for SEMs: the number of indicators and the magnitude of factor loadings. Following the two-indicator rule (Bollen, 1989), the minimum value for the number of indicators was set to 2, and the maximum value was selected as 4, generating three numbers of indicators (i.e., two, three, and four). For the factor loading ( $\lambda$ ), 0.4 was chosen as a minimum suggested value from the literature (e.g., Ford et al., 1986; Wang & Wang, 2012), and 0.7 was selected to ensure sufficient convergent validity (Kline, 2016). The two middle values (0.5 and 0.6) were also included, yielding four different factor loading conditions. Under each factor loading condition, all  $\lambda$ s were set as equal. Because all variables, including indicators, were assumed to have a variance of 1 in this study, the measurement error variances for the indicators were calculated as  $1 - \lambda^2$ .

In sum, a total of  $3 \times 2 \times 4 \times 2 = 48$  (i.e., three levels of indirect effect sizes, two levels of direct effect, four mediation models, and two testing methods) conditions were considered in path models, while a total of  $3 \times 2 \times 4 \times 3 \times 4 \times 2 = 576$  (i.e., three levels of indirect effect sizes, two levels of direct effect, four mediation models, three numbers of indicators, four magnitudes of factor loadings, and two testing methods) conditions were manipulated in SEMs. In addition, the impact of missing data on sample size requirements was examined by including five proportions of missing data: 0%, 5%, 10%, 15%, and 20%. The range of 0% to 20% was chosen to resemble the values used in Wolf et al. (2013). The missing data were generated as missing completely at random (Little & Rubin, 1989). The missing conditions were not fully crossed with

the other conditions explained earlier. Using the delta and bootstrap methods, the four SEMs with the partial mediation condition were considered only under the fixed condition of three indicators, 0.7 factor loading, and medium effect size.

One thousand replications were generated for each condition. For the bootstrap method, the number of bootstrap samples was chosen to be 1,000 following previous studies (Fritz et al., 2012; Kim, 2012; Shrout & Bolger, 2002; Tofighi & Kelley, 2020). In order to determine a sample size necessary for each condition, sample sizes from 10 to the minimum required sample size, in increments of 10, were considered. The minimum required sample size was capped at 10,000 in consideration of potential sampling difficulty in real applications.

### Data Analysis

All data generation and Monte Carlo simulations were carried out using *Mplus* 8.3 (L. K. Muthén & Muthén, 1998-2020)<sup>4</sup> and Python (Van Rossum & Drake, 1995).<sup>5</sup> In order to test mediation effects, two testing methods were used: the percentile bootstrap method and the multivariate delta method.<sup>6</sup> The delta method was chosen because it has been used frequently in recent substantive research (e.g., Heatly & Votruba-Drzal, 2017; Wade et al., 2018; Wentzel et al., 2018) and is easily implemented in *Mplus* by default. The percentile bootstrap method was also considered because it has shown to be more powerful than the delta method (i.e., Sobel test), have less inflated Type I error, and offer better coverage than the bias-corrected bootstrap test (Hayes & Scharkow, 2013). Also, the percentile bootstrap tended to show low Type I error rates and high coverage rates, compared to the bias-corrected bootstrap test and the bias-corrected and accelerated bootstrap test (Falk & Biesanz, 2015). Hence, the percentile bootstrap method as well as the delta method were used to test the significance of the mediation effect for each generated data set.

Considering the sample size study of L. K. Muthén and Muthén (2002), three criteria were considered to determine the required minimum sample sizes in the present study: parameter bias, 95% coverage, and power.<sup>7</sup> Parameter bias is the difference between a population parameter value and the average of parameter estimates over the replications of a Monte Carlo study. Ninety-five percent coverage is the proportion of replications for which the 95% confidence interval contains the true parameter value. Power is the proportion of replications for which a null hypothesis is correctly rejected for each nonzero parameter at an  $\alpha$  level of .05. The first criterion to determine sample sizes is that the parameter bias does not exceed 10% for any parameter whose population value is not equal to zero in the model. The second criterion is that 95% coverage remains between .91 and .98, and is applied to all parameters in a model being considered. Other suggestions, such as 95% coverage between .925 and .975 (Algina et al., 2005), were also proposed, but the present study followed L. K. Muthén and Muthén (2002). The third criterion is that the sample size is chosen to maintain power close to .8 or greater (Cohen, 1988; Fritz & MacKinnon, 2007) with

**Table 3.** Minimum Required Sample Sizes for Path Models.

Method	Effect size	Partial mediation				Complete mediation			
		Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Delta	Small	680	1,170	1,610	1,090	670	1,000	1,440	1,040
	Medium	100	150	220	150	100	130	200	140
	Large	50	60	90	50	50	50	80	50
	Average	277	460	640	430	273	393	573	410
Bootstrap	Small	560	980	1,210	920	560	830	1,090	880
	Medium	80	130	160	120	90	110	150	110
	Large	40	50	70	40	60	50	60	40
	Average	227	387	480	360	237	330	433	343

Note. Effect size: The effect size of the indirect effect(s).

an alpha level of .05. The power is evaluated only for specific parameter effects of interest, that is, mediation effects in this study.

## Results

For Models 1 to 4, minimum required sample sizes that met the three criteria (parameter bias, 95% coverage, and power) were obtained under each simulation condition. Before summarizing simulation results, the overall patterns regarding the three criteria are briefly described. In path models, once the power criterion was met, the other two criteria were already satisfied in most conditions. In other words, it was relatively difficult to achieve the power of .8 for detecting mediation effects. Such a pattern was again observed with SEMs. In addition, large parameter biases frequently occurred with the conditions of two indicators under Model 1 and the conditions of large indirect effect size under Models 1, 2, and 4, especially with the bootstrap method. Finally, nonconvergence rates were also checked: They ranged from 0% to 5% in most conditions, which did not seem very serious based on the results from previous simulation research (e.g., Gagné & Hancock, 2006; Jia et al., 2014; Moineddin et al., 2007). The nonconvergence issue occurred with the two-indicator conditions and mostly under Models 1 and 4 with some complete mediation conditions.

### Path Models

The minimum required sample sizes of the path mediation models are provided in Table 3. The minimum required sample sizes ranged from 50 to 1,610 for the delta method, while they ranged from 40 to 1,210 for the bootstrap method. The largest sample sizes in both methods were obtained in the partial mediation condition of Model 3 with the small effect size. Overall, Model 3 required the largest sample size

on average ( $\bar{n} = 531.7$ ), which was larger than the required sample sizes obtained for Model 2 ( $\bar{n} = 392.5$ ) and Model 4 ( $\bar{n} = 385.8$ ). Model 1, which is the simplest of the four models, required the smallest sample size on average ( $\bar{n} = 253.3$ ). It is well-known that as the complexity of the model increases, the sample size required for accurate and stable estimation is expected to increase. In our case, Model 3 can be considered to be the most complex model, because it has a multiple-step mediation in addition to two single-step mediation effects, whereas Models 2 and 4 included only two single-step mediation effects in each model. The partial mediation conditions required larger sample sizes on average than the complete mediation conditions. This was probably attributable to the fact that one or two fewer parameters were estimated in the complete mediation conditions due to the constraint of  $c' = 0$ , depending on the model being considered. The difference in the required sample sizes between the partial and complete mediation conditions ranged from  $-20$  to  $170$  across conditions with an average of  $33$ . Larger differences tended to be observed as the effect size became smaller.

Regardless of the models, as the level of indirect effect size decreased, the required sample size increased. In particular, when the effect size decreased from the medium to small conditions, the sample sizes dramatically increased compared to the change from the large to medium conditions. The required sample sizes increased more than six times (up to eight times) when the effect size changed from the medium to small conditions. However, the required sample sizes changed about two or three times from the large to medium effect size conditions.

Last, the bootstrap method required smaller sample sizes than the delta method in most conditions. In particular, as the indirect effect size decreased and the model complexity increased, the difference in the required sample sizes between the two methods became larger. For example, the largest difference was observed in Model 3 with the small effect size and the partial mediation condition. In this case, the bootstrap method required  $1,210$ , while the delta method needed  $1,610$ . Overall, the differences in the required sample sizes between the two methods ranged from  $110$  to  $400$  under the small effect size conditions, while they ranged from  $10$  to  $60$  under the medium effect size conditions. More comparisons between the two methods are provided later in Table 6.

### *Structural Equation Models*

The minimum required sample sizes of the structural equation mediation models with partial and complete mediations are provided in Table 4 and Table 5, respectively. The smallest sample size ( $n = 70$ ) was required for the complete mediation condition of Model 1 with the large indirect effect size, four indicators, and  $0.7$  loadings in the bootstrap method. As indicated earlier, the required sample sizes were capped at  $10,000$ , and some conditions required sample sizes over  $10,000$  ( $n > 10,000$ ) for accurate estimation. Those conditions were all associated with the small effect size and mostly with the conditions of two indicators and  $0.4$  loadings. Like the results of

**Table 4.** Minimum Required Sample Sizes for Structural Equation Models with Partial Mediation.

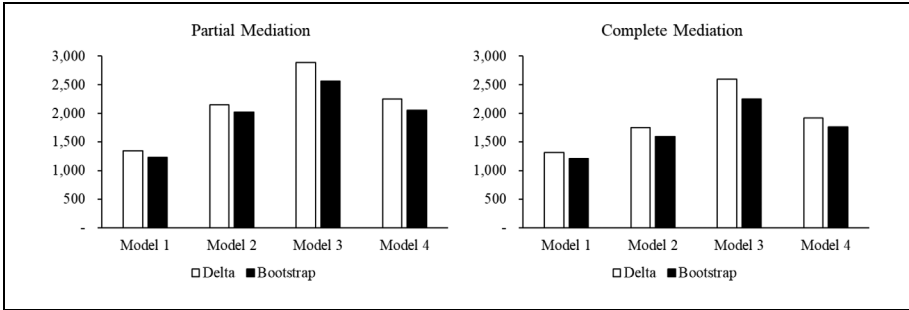
Method	Effect size	Standardized factor loading	Model 1				Model 2				Model 3				Model 4			
			Number of indicators				Number of indicators				Number of indicators				Number of indicators			
			2	3	4		2	3	4		2	3	4		2	3	4	
Delta	Small	0.4	9,410	4,890	3,330	> 10,000	8,950	6,500	> 10,000	> 10,000	> 10,000	8,420	> 10,000	9,350	6,390			
		0.5	4,440	2,590	1,940	8,010	4,720	3,670	> 10,000	> 10,000	6,560	4,800	7,840	4,890	3,590			
		0.6	2,480	1,650	1,340	4,470	2,900	2,460	5,820	4,140	3,280	4,480	3,000	2,390				
	Medium	0.7	1,590	1,180	1,020	2,870	2,130	1,850	3,830	2,950	2,480	2,830	2,080	1,810				
		0.4	1,940	930	650	2,830	1,420	1,030	4,340	2,300	1,520	3,150	1,600	1,100				
		0.5	850	480	360	1,260	740	570	1,950	1,110	820	1,370	770	570				
		0.6	470	290	240	680	450	380	1,000	650	510	730	470	360				
Bootstrap	Small	0.7	290	190	170	420	300	280	610	450	380	430	320	260				
		0.4	1,950	880	590	2,790	1,330	700	5,440	2,450	1,520	4,040	1,820	1,200				
		0.5	710	360	240	1,240	510	330	2,110	1,010	660	1,350	730	460				
	Medium	0.6	330	180	140	480	260	180	920	510	350	550	370	200				
		0.7	170	120	100	210	160	120	440	270	220	250	190	120				
		0.4	8,870	4,420	2,920	> 10,000	8,100	5,860	> 10,000	> 10,000	9,670	6,590	> 10,000	8,270	5,360			
		0.5	4,220	2,290	1,660	8,000	4,160	3,190	9,240	5,150	3,770	7,190	4,170	3,050				
Large	0.6	2,410	1,410	1,140	4,470	2,550	2,230	5,190	3,160	2,510	4,060	2,560	2,040					
	0.7	1,590	960	870	3,010	1,840	1,670	3,430	2,240	1,910	2,630	1,800	1,520					
	0.4	1,650	870	620	2,920	1,490	1,050	4,010	2,090	1,320	2,810	1,480	1,010					
	0.5	760	450	330	1,280	720	550	1,730	980	710	1,270	700	520					
	0.6	420	260	210	700	410	340	900	570	430	700	430	320					
	0.7	270	180	150	420	280	250	550	370	310	410	280	220					
	0.4	1,730	880	590	2,510	1,140	650	5,200	2,440	1,450	3,990	1,810	1,200					
Large	0.5	600	340	230	900	450	290	2,010	990	650	1,330	700	460					
	0.6	290	180	140	350	260	170	880	480	340	510	370	200					
	0.7	160	110	90	200	160	110	420	250	200	230	190	110					

Note. Effect size: The effect size of the indirect effect(s).

**Table 5.** Minimum Required Sample Sizes for Structural Equation Models with Complete Mediation.

Method	Effect size	Standardized factor loading	Model 1				Model 2				Model 3				Model 4			
			Number of indicators				Number of indicators				Number of indicators				Number of indicators			
			2	3	4		2	3	4		2	3	4		2	3	4	
Delta	Small	0.4	> 10,000	5,010	3,450	> 10,000	7,580	5,340	> 10,000	> 10,000	10,000	7,450	> 10,000	8,930	6,000			
		0.5	5,060	2,670	1,990	6,850	3,890	3,070	9,510	5,830	4,320	7,480	4,580	3,440				
		0.6	2,840	1,710	1,350	3,820	2,450	2,090	5,310	3,730	2,930	4,230	2,840	2,280				
	Medium	0.7	1,770	1,200	1,060	2,480	1,780	1,600	3,460	2,630	2,220	2,670	2,000	1,740				
		0.4	1,650	780	530	2,100	1,090	790	3,950	2,050	1,400	2,390	1,230	870				
		0.5	780	410	320	940	570	440	1,730	990	720	1,090	630	480				
		0.6	430	250	210	550	350	290	910	580	440	600	400	310				
Bootstrap	Small	0.7	270	180	160	340	250	230	540	400	330	370	270	230				
		0.4	930	430	290	1,240	600	400	3,930	1,800	1,130	1,190	580	390				
		0.5	420	220	160	530	290	210	1,530	760	520	510	280	200				
	Medium	0.6	220	130	110	270	170	130	700	380	290	290	170	130				
		0.7	160	90	80	160	110	100	370	230	180	160	110	90				
		0.4	> 10,000	4,140	2,830	> 10,000	6,430	4,450	> 10,000	8,580	5,840	> 10,000	8,020	5,320				
		0.5	4,960	2,240	1,650	6,300	3,360	2,640	8,150	4,480	3,280	6,730	4,000	2,940				
Large	0.6	3,070	1,390	1,120	3,630	2,120	1,770	4,660	2,810	2,230	3,800	2,470	2,000					
	0.7	1,880	960	850	2,400	1,520	1,350	3,040	2,010	1,680	2,370	1,730	1,480					
	0.4	1,300	700	510	1,870	1,020	720	3,450	1,780	1,200	2,240	1,190	830					
	0.5	730	360	280	880	530	410	1,500	850	600	1,030	600	450					
	0.6	450	230	190	520	320	280	800	500	390	580	370	280					
	0.7	260	160	140	330	220	200	490	330	270	360	240	200					
	0.4	820	390	270	1,220	570	400	4,190	1,910	1,140	1,190	590	400					
Large	0.5	330	190	150	460	270	210	1,620	760	510	480	320	210					
	0.6	230	120	100	250	160	130	710	380	280	250	170	130					
	0.7	130	90	70	150	110	100	360	220	170	150	110	90					

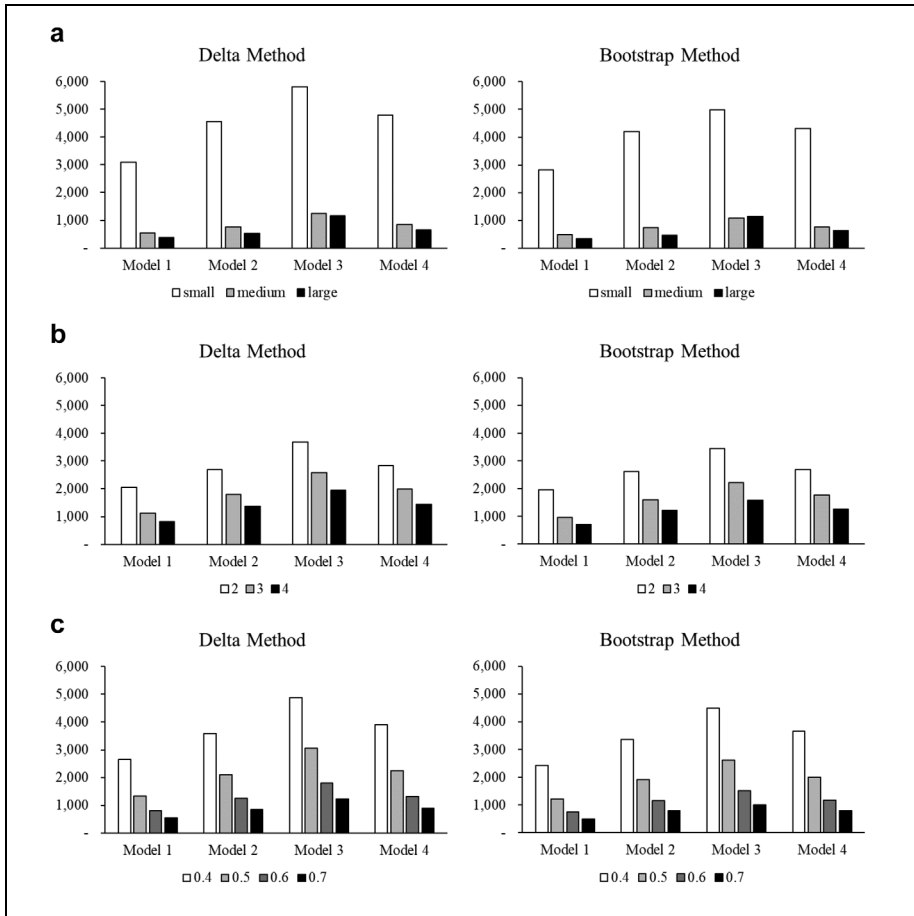
Note. Effect size: The effect size of the indirect effect(s).



**Figure 2.** Average required sample sizes for structural equation models.

the path models, Model 3 required larger sample sizes on average ( $\bar{n}=2,572$ ) than Model 4 ( $\bar{n}=1,996$ ), which required larger sample sizes than Model 2 ( $\bar{n}=1,877$ ). Model 1 showed the smallest required sample sizes ( $\bar{n}=1,273$ ) as expected. This pattern is clearly shown in Figure 2, which displays the required sample sizes across the models. These were separately averaged for each model with each testing method in the partial and complete mediation conditions. The partial mediation conditions needed larger average sample sizes than the complete mediation conditions. In particular, Models 2 to 4 with the partial mediation required much larger sample sizes (about 340 more) than those with the complete mediation. However, for Model 1, the partial mediation condition required just a little larger sample sizes (about 30 more) than the complete mediation condition on average. In fact, some complete mediation conditions in Model 1 based on Tables 4 and 5 required larger sample sizes than the partial mediation conditions. All the small effect size conditions in the delta method and some two-indicator conditions in the bootstrap method showed such a pattern. This was attributed to the fact that parameter biases were more serious in the two-indicator and/or small effect size conditions with the complete mediation than those with the partial mediation in Model 1. Therefore, a caution would be necessary when the complete mediation of Model 1 is used with two indicators and/or small effect sizes. Figure 2 also shows that the delta method required larger sample sizes than the bootstrap method across all models on average. There were some conditions (19 out of 288 conditions) that showed less sample sizes in the delta method compared to the bootstrap method (see Tables 4 and 5). However, no systematic pattern regarding those 19 conditions was observed. Like the path models, the largest difference between the two methods was found in Model 3. The delta method required about 330 larger sample sizes than the bootstrap method on average in Model 3. Model 4 and Model 2 showed 143 and 181 differences between the two methods, respectively. The smallest difference was about 115, which was observed in Model 1.

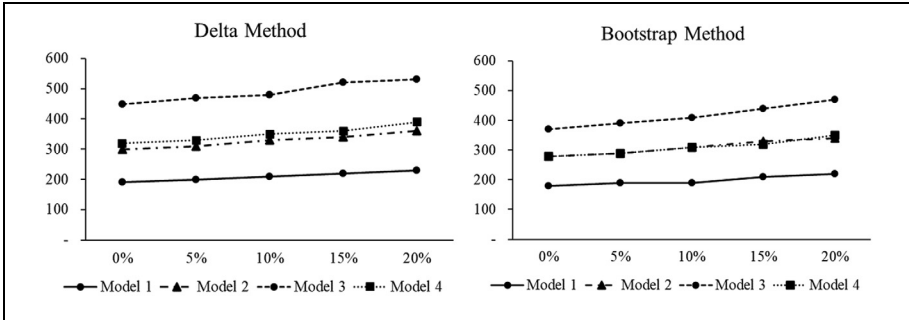
Regarding the results in Tables 4 and 5, the influences of three simulation factors (i.e., the level of indirect effect sizes, the number of indicators, and the magnitude of factor loadings) are illustrated in Figure 3a through Figure 3c. Figure 3a shows the



**Figure 3.** The impact of (a) indirect effect sizes, (b) number of indicators, and (c) magnitude of factor loadings on required sample sizes for structural equation models.

average minimum required sample sizes for Models 1 to 4 across the three indirect effect sizes in the delta method (left) and the bootstrap method (right). The partial and complete mediation conditions were averaged in each figure. In general, the required sample sizes increased as the indirect effect size decreased. This tendency was dramatic when the effect size moved from the small to medium conditions. The required sample sizes in the small effect size conditions were, on average about 5.5 times larger than those in the medium effect size conditions. However, the required sample sizes were raised 1.3 times on average when the effect size changed from the large to medium conditions. Meanwhile, there were some unexpected deviations when the effect size moved from the large to medium conditions. That is, for some





**Figure 4.** The impact of missing proportions on required sample sizes for partially mediated structural equation models.

simulated conditions, larger sample sizes were required with the large effect size than with the medium effect size. This pattern was observed only in the partial mediation conditions. An examination of the results in Table 4 reveals that this deviated pattern (13 out of 288 conditions) was observed in Models 1, 3, and 4 when the number of indicators was small (i.e., two indicators) and/or the magnitude of factor loadings was small (i.e., 0.4). This pattern disappeared as the number of indicators and/or the size of factor loadings increased. Figure 3b displays the effect of the number of indicators on the minimum required sample sizes for each model. The number of indicators is relatively easier to control in reality by researchers. Thus, it would be valuable to examine the effect of the number of indicators in sample size studies. It is evident from the results that the sample sizes decreased as the number of indicators increased across all conditions. This tells us that more indicators (i.e., more parameters) did not always make the model more complex. As shown in Figure 3b, the decreasing rates of the sample sizes across the conditions of the number of indicators were slightly larger from two to three indicators than from three to four indicators. Last, Figure 3c illustrates the required sample sizes across four different factor loadings in each model. Overall, the required sample sizes decreased as the magnitude of factor loadings increased across all models. Like the effect of the number of indicators, the dropping rates declined as the factor loadings increased. There was no reverse pattern associated with the effect of the number of indicators or the effect of factor loadings given the other factors stayed constant (see also Tables 4 and 5). Overall, the two testing methods showed similar patterns regarding the three simulation factors, but the bootstrap method required less sample sizes than the delta method as mentioned earlier.

The impact of missing data on sample size requirements was examined for the four SEMs with partial mediation by including five proportions of missing data: 0%, 5%, 10%, 15%, and 20%. As explained above, the fixed condition of three indicators, 0.7 factor loading, and medium effect size was only considered. Figure 4 shows that higher proportions of missing data in the four models generally required larger

**Table 6.** Averaged Percentages Based on the Ratios of the Required Sample Sizes in the Bootstrap Method to Those in the Delta Method.

Model	Model 1	Model 2	Model 3	Model 4	Average
Path model	89.3%	86.9%	75.2%	81.3%	83.2%
SEM	90.9%	92.9%	89.2%	93.1%	91.5%
Average	90.1%	89.9%	82.2%	87.2%	87.4%

Note. The numbers were calculated as the ratios of the required sample sizes in the bootstrap method to those in the delta method multiplied by 100%, which were then averaged across conditions. SEM = structural equation model.

sample sizes, as found in previous studies (e.g., Wolf et al., 2013). For example, Model 1 with the delta method required a minimum sample size of 190 when there were no missing data. However, the sample size increased to 230 with 20% missing data. Overall, models with 20% missing data necessitated, on average, about 20% and 24% increases in sample size requirements for the delta method and the bootstrap method, respectively. Among the three criteria (parameter bias, 95% coverage, and power), it was the most difficult to achieve the power of .8 for detecting indirect effects across all conditions. Although the increasing patterns were similar among the four models and the two testing methods in Figure 4, the increasing rates tended to get higher as the model complexity increased (i.e., Model 3 > Model 4, Model 2 > Model 1).

### Comparisons of the Findings

Based on the results of the path models and the SEMs, the averaged required sample size for each path model in the delta method ranged from 253.3 to 531.7, while those in the SEMs ranged from 1,273 to 2,572. The inclusion of latent variables with indicators required much larger sample sizes than the corresponding path models for accurate estimation of parameters. For example, for the simple model (Model 1) with the partial mediation and the bootstrap method, the SEMs needed 1,230 on average, which is 5.4 times larger than the path models that required 227 on average. Across all four models, the SEMs required five times the sample sizes required for the corresponding path models. Therefore, if it is not feasible to obtain sufficient sample sizes for mediation models, using path models would be a plausible option to obtain an accurate and stable estimation of model parameters.<sup>8</sup>

The required sample sizes for the path models and SEMs through the Monte Carlo study can be compared with the search results in Table 1. If the median values were considered from Table 1, quite smaller sample sizes were actually being used in applied psychological studies than the minimum required sample sizes from the current study. Overall, this tendency was more severe in the SEMs, though the path models also showed quite large discrepancies between the survey results and the

simulation results. This comparison implies that applied researchers, in reality, tended to use insufficient sample sizes to estimate their models of interest, which might result in unstable and inaccurate estimation of the model parameters including mediation effects, possibly leading to invalid inferences concerning their research questions.

Finally, based on the results from the delta and bootstrap methods across conditions, Table 6 displays average percentages. These numbers were calculated as the ratios of the required sample sizes in the bootstrap to those in the delta multiplied by 100%, which were then averaged across conditions. The bootstrap method required less sample sizes (i.e., more powerful) than the delta method. The effect of using the bootstrap instead of the delta was larger in path models than SEMs. In path models, the bootstrap method, on average, required 83.2% of the sample size needed in the delta method. Among the four models, Model 3 showed the largest reduction in the sample size (i.e., about 25%) by using the bootstrap. In SEMs, the bootstrap method, on average, required 91.5% of the sample size needed in the delta method. Unlike the path models, all four SEMs tended to show similar results.

## **Discussion and Conclusion**

The primary purpose of the present study was to provide practical and general guidelines on minimum sample size requirements for accurate estimation of four mediation models, which are commonly used in the literature. That is, the aim was to provide substantive researchers with concrete and tangible guidance on the required sample sizes for frequently used mediation models. We compared SEMs with path models to investigate the effect of including latent variables with indicators on sample size requirements through a Monte Carlo simulation study. The overall results of this study support the following conclusions. First, the actual sample sizes used in the literature are much less than the minimum required sample sizes identified in the current Monte Carlo study. Second, the sample size requirements are affected by several factors, such as the model complexity, the level of indirect effect sizes, the number of indicators, and the magnitude of factor loadings. As the model complexity increased, the required sample sizes increased. Also, as the level of indirect effect sizes, the number of indicators, and the magnitude of factor loadings decreased, the required sample sizes increased. Testing methods also affected the results. The bootstrap method required less sample sizes than the delta method for the similar level of accuracy in the parameter estimation. By using the bootstrap method instead of the delta method, more reduction in sample sizes was obtained in the path models, compared to the SEMs. In addition, the missing proportion had an effect on the minimum required sample sizes. The sample sizes for the four SEMs increased as the missing proportion increased.

Some important findings and relevant suggestions are as follows. First, regarding the influence of the indirect effect size on the sample size, the dropping rates in sample sizes were much higher for the effect size conditions from small to medium than

for the conditions from medium to large. A similar pattern was also observed in previous studies (e.g., Liu & Wang, 2019; Wolf et al., 2013). This implies that researchers need to choose predictors (and/or mediators) with at least medium effect sizes, if ever possible. In that case, a much smaller sample size would be required than using predictors with small effect sizes. From a different point of view, when a researcher determines a sample size for his/her own research, he/she may assume a medium effect size rather than small or large effect sizes, as Cohen (1988) suggested that a medium effect size should represent the average effect for a given research area. Another finding in the simulations was that the level of indirect effect sizes seemed to interact with the number of indicators and the magnitude of factor loadings in affecting sample size requirements. The large effect size conditions required larger sample sizes than the medium effect size conditions, particularly when the number of indicators was small (i.e., two indicators) and/or the factor loading was small (i.e., 0.4). These reverse patterns disappeared as the number of indicators and/or the magnitude of the factor loadings increased. Previous studies (e.g., Liu & Wang, 2019; Wolf et al., 2013) also showed similar results (i.e., sample sizes increased as indirect effect sizes increased).

Second, the number of indicators also played an important role in estimating the mediation model of interest. The required sample sizes decreased as the number of indicators increased. Like the influence of the effect size, the dropping rates were higher for the conditions from two to three indicators than for the conditions from three to four indicators. Also, several unexpected results were found regarding the two-indicator conditions as explained earlier. It has been known that at least two indicators per factor are needed when more than one factor is in the model (Bollen, 1989). However, when the number of indicators was two, some reverse patterns were observed as previously mentioned. Marsh et al. (1998) also suggested not to use two indicators in the context of CFA. Instead, they mentioned that researchers are typically recommended to use three or more indicators per factor. Based on the findings in the present study, we also suggest researchers use at least three indicators, if possible, to obtain both accurate estimation and affordable sample sizes. Researchers also need to keep in mind that according to the current simulation results, increasing the number of indicators (at least up to four) did not make the model more complex.

Third, regarding the effect of factor loadings (0.4 to 0.7), when the factor loading was 0.4, the largest sample size was needed. As the factor loadings increased, the required sample sizes decreased. A similar pattern was also found in Wolf et al. (2013), who studied CFA with three magnitudes of factor loadings (0.5, 0.65, 0.8). In the current study, the dropping rate in sample sizes was the largest when the factor loadings moved from 0.4 to 0.5. Wang and Wang (2012) suggested that standardized factor loadings need to be at least 0.4. Still, as shown in the simulation results, having 0.4 loadings for all indicators can hinder stable estimation, eventually leading to large sample sizes required. Therefore, we suggest that researchers use more reliable indicators for a latent variable (e.g., more than 0.4 factor loading). If applicable,

researchers may acquire a reasonable conjecture on the range of factor loadings from previous findings in the literature.

Fourth, findings of this study also emphasize that researchers should pay attention to the existence of missing data in determining required samples sizes. Based on the results, it was necessary to have about 22% increase in sample size requirements, on average, under 20% missing data, compared to 0% missing data. Also, the missing data affected statistical power for detecting indirect effects. These results imply that ignoring missing data when determining sample sizes may result in unstable and inaccurate estimation of the model parameters including mediation effects.

When researchers decide sample sizes for their mediation models, they typically consider sample sizes used in previous studies or follow some popular rules-of-thumb, such as 5 or 10 observations per estimated parameter (Bentler & Chou, 1987; Bollen, 1989) or a minimum sample size of 100 or 200 (Boomsma, 1982, 1985). The former approach would be problematic because some previous studies, as we see in Table 1, arbitrarily determined less than optimal sample sizes without careful investigations. The latter approach would also be improper because the rules-of-thumb are not model-specific and may lead to inordinately larger or smaller sample sizes than needed (Wolf et al., 2013). We hope that applied researchers will discover the findings in the present study applicable to their work and use them to determine minimum sample size requirements for their models of interest. It should be noted that the sample sizes in the result tables need to be considered as reference materials, not as absolute values. In particular, mediation models with small samples need to be used with caution. For example, using the required sample sizes as low as 40 for path models (see Table 3) is likely to violate model assumptions such as multivariate normality. In such cases, Bayesian estimation with informative priors may be used as suggested in the literature (e.g., Koopman et al., 2015; Miočević et al., 2017).

Although SEMs are known to be superior to path models because they take measurement errors into account, researchers can choose path mediation models instead of corresponding SEMs if insufficient sample sizes were inevitably collected (e.g., a study with a small-sized population). Also, in case of choosing the bootstrap method versus the delta method, researchers should choose the bootstrap method, which resulted in about 13% (17% for path models and 8.5% for SEMs) reduction on the required sample sizes compared to the delta method on average. If researchers have already gathered a sample of subjects for fitting a mediation SEM, they should then focus on the number of indicators, which is a relatively controllable factor by researchers. When the sample size is not large enough to get an accurate and stable estimation for SEMs, the researchers should increase the number of indicators per factor. If possible, researchers need to take other factors, such as indirect effect sizes and factor loadings, into consideration, as explained above. Furthermore, the selection of partial mediation versus complete mediation should be determined by theoretical backgrounds or actual (observed) relationships among the variables being considered.

This study has several limitations. First, although we examined the four mediation models which were frequently used in the literature, we did not consider more complex mediation models, such as multilevel mediation, longitudinal mediation, moderated mediation, and so on. These models have received recent attention (Chen et al., 2018; Hu et al., 2018; Stäbler et al., 2017; Wentzel et al., 2018), but including all these models was not possible in a single study. Second, we only evaluated SEMs with two, three, and four indicators. According to the results of the current study, as the investigation moved from two to three, and then from three to four indicators, the required sample sizes decreased. However, we did not investigate the further effect of more numbers of indicators (e.g., five, six, or seven). Hence, it might be interesting to examine the effect of more than four indicators on sample size requirements in the context of mediation analysis, as the number of indicators can be relatively easily manipulated by researchers, as compared to the other factors such as effect sizes or factor loadings. Third, the numbers reported in the current study were not sufficient required sample sizes for accurate estimation of the models but rather minimum required sample sizes. Because of that, even though the three criteria (parameter bias, 95% coverage, and power) were satisfied with a sample size of 500, for example, it was still possible that a part of those criteria may not be met with a sample of 510. Finally, the findings and suggestions are limited to the conditions included in this Monte Carlo study. For example, we did not consider nonnormal data and categorical (or ordinal) variables. It remains a task to examine systematically the impact of other aspects of research that were not included in this study.

The present study was carried out to provide applied researchers with practical guidelines in determining sample sizes when conducting analyses with some frequently used mediation models, including both path models and SEMs. Although the simulation approach limits the generalizability of the findings, they should be useful in understanding which factors affect sample sizes of mediation models and what considerations are involved in choosing sample sizes for mediation analysis.

## Appendix A

The coefficient,  $a$ , for example, was calculated as follows. In the simple regression (without intercept),

$$Y = aX \rightarrow a = \frac{Y}{X}.$$

By standardizing all components above,

$$\frac{Y}{S_Y} = a^s \frac{X}{S_X} \rightarrow a^s = \frac{Y S_X}{X S_Y} = a \frac{S_X}{S_Y},$$

where  $a^s$  is the standardized coefficient.

By squaring both sides,

$$(a^s)^2 = a^2 \frac{Var(X)}{Var(Y)},$$

$$(a^s)^2 = r_{XY}^2 = R^2.$$

Because  $X$  and  $Y$  were standardized in this study,

$$a^2 = R^2,$$

$$\therefore a = \sqrt{R^2}.$$

When the effect size was small (0.02),  $a = \sqrt{0.02} = 0.14$ . These calculations were also applied to  $b$  and  $c'$ . To calculate the error variances of  $d_M$  and  $d_Y$ , the covariance between  $X$  and  $M$  was needed and derived as follows:

$$r_{XM} = \frac{Cov(X, M)}{S_X S_M},$$

$$R^2 = r_{XM}^2 = \frac{Cov(X, M)^2}{Var(X)Var(M)},$$

$X$  and  $M$  were standardized in this study, and thus,  $Cov(X, M) = \sqrt{R^2}$ . When the effect size was small (0.02),  $Cov(X, M) = \sqrt{0.02} = 0.14$ . Finally, the error variances of  $d_M$  and  $d_Y$  were calculated in the following way. When  $a$ ,  $b$ , and  $c$  were all 0.14 (small effect size) as in Table 1,

$$M = aX + d_M,$$

$$Var(M) = a^2 Var(X) + Var(d_M),$$

$$1 = 0.02 \cdot 1 + Var(d_M),$$

$$\therefore Var(d_M) = 0.98.$$

$$Y = c'X + bM + d_Y,$$

$$Var(Y) = c'^2 Var(X) + b^2 Var(M) + 2c'b Cov(X, M) + Var(d_Y),$$

$$1 = 0.02 \cdot 1 + 0.02 \cdot 1 + 2\sqrt{0.02}\sqrt{0.02}\sqrt{0.02} + Var(d_Y),$$

$$\therefore Var(d_Y) = 0.95.$$

## Appendix B

Monte Carlo simulation *Mplus* codes for three example conditions are provided below.

1. Syntax for complete mediation path Model 1 with delta method and medium effect size.

## MONTECARLO:

```

NAMES ARE x m y;           !Define variable names
NOBS = 100;                !Power is estimated for this sample size
NREPS = 1000;             !Number of replications
SEED = 560910;           !Seed value for random number generator

```

## ANALYSIS:

```

ESTIMATOR = ML;
TYPE = general;

```

## MODEL POPULATION:

```

[X @0];                   !Factor means to 0
[M @0];                   !Intercept to 0
[Y @0];

X @1;                     !Total variance of X
M @0.87;                  !Residual variance of M
Y @0.87;                  !Residual variance of Y

M ON X *0.36;             !a path
Y ON M *0.36;             !b path
Y ON X @0;                !c' path

```

## MODEL:

```

[X @0];                   !Factor means to 0
[M @0];                   !Intercept to 0
[Y @0];

X @1;                     !Total variance of X
M @0.87;                  !Residual variance of M
Y @0.87;                  !Residual variance of Y

M ON X *0.36;             !a path
Y ON M *0.36;             !b path
Y ON X @0;                !c' path

```

## MODEL INDIRECT:

```

y IND x;                  !To obtain mediated effect

```

## OUTPUT : TECH9;



2. Syntax for partial mediation SEM 1 with bootstrap method, medium effect size, three indicators, and 0.7 factor loadings.

MONTECARLO:

NAMES ARE x1-x3 m1-m3 y1-y3; !Define variable names  
 NOBS = 180; !Power is estimated for this sample size  
 NREPS = 1000; !Number of replications  
 SEED = 560910; !Seed value for random number generator

ANALYSIS:

TYPE = general; BOOTSTRAP = 1000;

MODEL POPULATION:

[x1-x3 @0]; !Intercept of indicators to 0  
 [m1-m3 @0];  
 [y1-y3 @0];  
 X BY x1-x3 \*0.7; !Measurement model, factor loading  
 M BY m1-m3 \*0.7;  
 Y BY y1-y3 \*0.7;  
 x1-x3 \*0.51; !Residual variance of indicators  
 m1-m3 \*0.51;  
 y1-y3 \*0.51;  
 [X @0]; !Factor means to 0  
 [M @0]; !Intercept to 0  
 [Y @0];  
 X @1; !Total variance of X  
 M @0.87; !Residual variance of M  
 Y @0.81; !Residual variance of Y  
 M ON X \*0.36; !a path  
 Y ON M \*0.36; !b path  
 Y ON X \*0.14; !c' path

MODEL:

[x1-x3 @0]; !Intercept of indicators to 0  
 [m1-m3 @0];  
 [y1-y3 @0];  
 X BY x1-x3 \*0.7; !Measurement model, factor loading  
 M BY m1-m3 \*0.7;  
 Y BY y1-y3 \*0.7;  
 x1-x3 \*0.51; !Residual variance of indicators  
 m1-m3 \*0.51;

```

y1-y3 *0.51;

[X @0];           !Factor means to 0
[M @0];           !Intercept to 0
[Y @0];

X @1;             !Total variance of X
M @0.87;         !Residual variance of M
Y @0.81;         !Residual variance of Y

M ON X *0.36;    !a path
Y ON M *0.36;    !b path
Y ON X *0.14;    !c' path

MODEL INDIRECT:  !To obtain mediated effect
  y IND x;

OUTPUT : CINTERVAL(BOOTSTRAP); TECH9;

```

3. Syntax for partial mediation SEM 1 with bootstrap method, medium effect size, 3 indicators, 0.7 factor loadings, and 10% missing.

MONTECARLO:

```

NAMES ARE x1-x3 m1-m3 y1-y3; !Define variable names
NOBS = 190;                   !Power is estimated for this sample size
NREPS = 1000;                 !Number of replications
SEED = 560910;                !Seed value for random number generator

!Missing data patterns
PATMISS = x1(0.1) x2(0.1) x3(0.1)
          m1(0.1) m2(0.1) m3(0.1)
          y1(0.1) y2(0.1) y3(0.1);
PATPROBS = 1.0;

```

ANALYSIS:

```
TYPE = general; BOOTSTRAP = 1000;
```

MODEL POPULATION:

```

[x1-x3 @0];           !Intercept of indicators to 0
[m1-m3 @0];
[y1-y3 @0];

X BY x1-x3 *0.7;      !Measurement model, factor loading
M BY m1-m3 *0.7;
Y BY y1-y3 *0.7;

x1-x3 *0.51;         !Residual variance of indicators
m1-m3 *0.51;

```

```

y1-y3 *0.51;

[X @0];           !Factor means to 0
[M @0];           !Intercept to 0
[Y @0];

X @1;             !Total variance of X
M @0.87;          !Residual variance of M
Y @0.81;          !Residual variance of Y

M ON X *0.36;     !a path
Y ON M *0.36;     !b path
Y ON X *0.14;     !c' path

MODEL:
[x1-x3 @0];       !Intercept of indicators to 0
[m1-m3 @0];
[y1-y3 @0];

X BY x1-x3 *0.7;  !Measurement model, factor loading
M BY m1-m3 *0.7;
Y BY y1-y3 *0.7;

x1-x3 *0.51;      !Residual variance of indicators
m1-m3 *0.51;
y1-y3 *0.51;

[X @0];           !Factor means to 0
[M @0];           !Intercept to 0
[Y @0];

X @1;             !Total variance of X
M @0.87;          !Residual variance of M
Y @0.81;          !Residual variance of Y

M ON X *0.36;     !a path
Y ON M *0.36;     !b path
Y ON X *0.14;     !c' path

MODEL INDIRECT:   !To obtain mediated effect
  y IND x;

```

OUTPUT : CINTERVAL(BOOTSTRAP); TECH9;


### Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A03041362).

## ORCID iDs

Mikyung Sim  <https://orcid.org/0000-0001-7918-644X>

Youngsuk Suh  <https://orcid.org/0000-0001-7197-6383>

## Notes

1. A “simple” mediation model indicates a mediation model with a single mediator (i.e., a model with one independent variable, one dependent variable, and one mediator).
2. The distinction between the complete and partial mediation was argued by Hayes (2018) for several conceptual reasons. However, these two modeling specifications are considered in this study because these terms are still abundant in the mediation literature.
3.  $c'$  in the complete mediation condition was fixed to 0, meaning that it was not estimated in the model being considered.
4. Some *Mplus* codes for Monte Carlo simulations are provided in Appendix B.
5. Python was used to perform automated simulations and summarization of the results using *Mplus*.
6. *Mplus* computes a test statistic for a mediation effect using the standard error defined by the multivariate delta method (MacKinnon, 2008).
7. L. K. Muthén and Muthén (2002) originally suggested four criteria (parameter bias, standard error bias, 95% coverage, and power). Among the four criteria, we did not consider the standard error bias, which is not a proper criterion for the bootstrap method.
8. A direct comparison between the two modeling approaches may not be fair because measurement errors are not considered in path models.

## References

- Algina, J., Keselman, H. J., & Penfield, R. D. (2005). Effect sizes and their intervals: The two-level repeated measures case. *Educational and Psychological Measurement, 65*(2), 241-258. <https://doi.org/10.1177/0013164404268675>
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78-117. <https://doi.org/10.1177/0049124187016001004>
- Boomsma, A. (1982). The 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* (pp. 149-285). North-Holland.
- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 50*, 229-242. <https://doi.org/10.1007/BF02294248>

- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley. <https://doi.org/10.1002/9781118619179>
- Chen, J., Bamberger, P. A., Song, Y., & Vashdi, D. R. (2018). The effects of team reflexivity on psychological well-being in manufacturing teams. *Journal of Applied Psychology, 103*(4), 443-462. <https://doi.org/10.1037/apl0000279>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
- Duckworth, A. L., White, R. E., Matteucci, A. J., Shearer, A., & Gross, J. J. (2016). A stitch in time: Strategic self-control in high school and college students. *Journal of Educational Psychology, 108*(3), 329-341. <https://doi.org/10.1037/edu0000062>
- Falk, C. F., & Biesanz, J. C. (2015). Inference and interval estimation methods for indirect effects with latent variable models. *Structural Equation Modeling, 22*(1), 24-38. <https://doi.org/10.1080/10705511.2014.935266>
- Foerstl, K., Franke, H., & Zimmermann, F. (2016). Mediation effects in the “purchasing and supply management (PSM) practice–performance link”: Findings from a meta-analytical structural equation model. *Journal of Purchasing & Supply Management, 22*(4), 351-366. <https://doi.org/10.1016/j.pursup.2016.08.001>
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*(2), 291-314. <https://doi.org/10.1111/j.1744-6570.1986.tb00583.x>
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*(3), 2330-239. <https://doi.org/10.1111/j.1467-9280.2007.01882.x>
- Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research, 47*(1), 61-87. <https://doi.org/10.1080/00273171.2012.640596>
- 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. [https://doi.org/10.1207/s15327906mbr4101\\_5](https://doi.org/10.1207/s15327906mbr4101_5)
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*(4), 408-420. <https://doi.org/10.1080/03637750903310360>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach*. Guilford Press.
- Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24*(10), 1918-1927. <https://doi.org/10.1177/0956797613480187>
- Heatly, M. C., & Votruba-Drzal, E. (2017). Parent- and teacher-child relationships and engagement at school entry: Mediating, interactive, and transactional associations across contexts. *Developmental Psychology, 53*(6), 1042-1062. <https://doi.org/10.1037/dev0000310>
- Hu, S., Hood, M., & Creed, P. A. (2018). Within-person relationship between career goal feedback and career-related stress: A weekly survey study. *Journal of Counseling Psychology, 65*(6), 681-689. <https://doi.org/10.1037/cou0000282>

- 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, 20*(1), 86-97. <https://doi.org/10.1080/10705511.2013.742388>
- Jia, F., Moore, W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (2014). Planned missing data designs with small sample sizes: How small is too small? *International Journal of Behavioral Development, 38*(5), 435-452. <https://doi.org/10.1177/0165025414531095>
- Kim, S.-Y. (2012). Sample size requirements in single- and multiphase growth mixture models: A Monte Carlo simulation study. *Structural Equation Modeling, 19*(3), 457-476. <https://doi.org/10.1080/10705511.2012.687672>
- Kline, R. B. (2016). *Principles and practices of structural equation modelling* (4th ed.). Guilford Press.
- Koopman, J., Howe, M., Hollenbeck, J. R., & Sin, H.-P. (2015). Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology, 100*(1), 194-202. <https://doi.org/10.1037/a0036635>
- Lachowicz, M. J., Preacher, K. J., & Kelley, K. (2018). A novel measure of effect size for mediation analysis. *Psychological Methods, 23*(2), 244-261. <https://doi.org/10.1037/met0000165>
- Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods and Research, 18*(2-3), 292-326. <https://doi.org/10.1177/0049124189018002004>
- Liu, X., & Wang, L. (2019). Sample size planning for detecting mediation effects: A power analysis procedure considering uncertainty in effect size estimates. *Multivariate Behavioral Research, 54*(6), 822-839. <https://doi.org/10.1080/00273171.2019.1593814>
- Lockman, J. D., & Servaty-Seib, H. L. (2018). Testing the predictions of the existential constructivist theory of suicide in a college student sample. *Journal of Counseling Psychology, 65*(3), 294-307. <https://doi.org/10.1037/cou0000278>
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Taylor & Francis. <https://doi.org/10.4324/9780203809556>
- Malone, J. C., Liu, S. R., Vaillant, G. E., Rentz, D. M., & Waldinger, R. J. (2016). Midlife Eriksonian psychosocial development: Setting the stage for late-life cognitive and emotional health. *Developmental Psychology, 52*(3), 496-508. <https://doi.org/10.1037/a0039875>
- 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. [https://doi.org/10.1207/s15327906mbr3302\\_1](https://doi.org/10.1207/s15327906mbr3302_1)
- Miočević, M., MacKinnon, D. P., & Levy, R. (2017). Power in Bayesian mediation analysis for small sample research. *Structural Equation Modeling, 24*(5), 666-683. <https://doi.org/10.1080/10705511.2017.1312407>
- Moineddin, R., Matheson, F. I., & Glazier, R. H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology, 7*, Article 34. <https://doi.org/10.1186/1471-2288-7-34>
- Muthén, L. K., & Muthén, B. O. (1998-2020). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- 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*(4), 599-620. [https://doi.org/10.1207/S15328007SEM0904\\_8](https://doi.org/10.1207/S15328007SEM0904_8)

- Muthén, B., & Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling, 22*(1), 12-23. <https://doi.org/10.1080/10705511.2014.935843>
- Ohrnberger, J., Fichera, E., & Sutton, M. (2017). The relationship between physical and mental health: A mediation analysis. *Social Science and Medicine, 195*, 42-49. <https://doi.org/10.1016/j.socscimed.2017.11.008>
- Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science, 8*(4), 379-386. <https://doi.org/10.1177/1948550617715068>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422-445. <https://doi.org/10.1037/1082-989X.7.4.422>
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290-312. <https://doi.org/10.2307/270723>
- Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish's achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic self-concept. *Journal of Educational Psychology, 109*(2), 191-207. <https://doi.org/10.1037/edu0000135>
- Tofighi, D., & Kelley, K. (2020). Indirect effects in sequential mediation models: Evaluating methods for hypothesis testing and confidence interval formation. *Multivariate Behavioral Research, 55*(2), 188-210. <https://doi.org/10.1080/00273171.2019.1618545>
- Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling, 17*(3), 510-534. <https://doi.org/10.1080/10705511.2010.489379>
- Van Rossum, G., & Drake, F. L., Jr. (1995). *Python reference manual*. Amsterdam.
- Wade, M., Madigan, S., Plamondon, A., Rodrigues, M., Browne, D., & Jenkins, J. M. (2018). Cumulative psychosocial risk, parental socialization, and child cognitive functioning: A longitudinal cascade model. *Developmental Psychology, 54*(6), 1038-1050. <https://doi.org/10.1037/dev0000493>
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus: Methods and applications*. Higher Education Press. <https://doi.org/10.1002/9781118356258>
- Wentzel, K. R., Muenks, K., McNeish, D., & Russell, S. (2018). Emotional support, social goals, and classroom behavior: A multilevel, multisite study. *Journal of Educational Psychology, 110*(5), 611-627. <https://doi.org/10.1037/edu0000239>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models. *Educational and Psychological Measurement, 73*(6), 913-934. <https://doi.org/10.1177/0013164413495237>
- Zhang, Z. (2014). Monte Carlo based statistical power analysis for mediation models: Methods and software. *Behavior Research Methods, 46*(4), 1184-1198. <https://doi.org/10.3758/s13428-013-0424-0>