Routledge
Taylor & Francis Group

Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: www.tandfonline.com/journals/hmbr20

A Comparison of Multilevel Mediation Modeling Methods: Recommendations for Applied Researchers

Christina K. Zigler & Feifei Ye

MULTIVARIATE BEHAVIORAL RESEARCH
2019, VOL. 54, NO. 3, 338–359
https://doi.org/10.1080/00273171.2018.1527676

A Comparison of Multilevel Mediation Modeling Methods: Recommendations for Applied Researchers

Christina K. Zigler[a,b] and Feifei Ye[c*]

[a]Department of Physical Medicine and Rehabilitation, School of Medicine, University of Pittsburgh, PA, USA; [b]Human Engineering Research Laboratories, VA Healthcare System, Pittsburgh, PA, USA; [c]School of Education, University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT
Multilevel structural equation modeling (MSEM) has been proposed as a valuable tool for estimating mediation in multilevel data and has known advantages over traditional multilevel modeling, including conflated and unconflated techniques (CMM & UMM). Recent methodological research has focused on comparing the three methods for 2-1-1 designs, but in regards to 1-1-1 mediation designs, there are significant gaps in the published literature that prevent applied researchers from making educated decisions regarding which model to employ in their own specific research design. A Monte Carlo study was performed to compare MSEM, UMM, and CMM on relative bias, confidence interval coverage, Type I Error, and power in a 1-1-1 model with random slopes under varying data conditions. Recommendations for applied researchers are discussed and an empirical example provides context for the three methods.

KEYWORDS
Mediation; Monte Carlo studies; multilevel modeling; structural equation modeling

Introduction

Mediation is an attractive methodology as it can shed light on complex relationships between two variables through a third "mediating" variable. This indirect pathway through the mediator helps to explain the underlying process through which a predictor affects an outcome (MacKinnon, 2008). There has been extensive research on estimating single level mediation for independent participants from simple random sampling (Hayes, 2013). However, multilevel data are commonly encountered in social science research, where children are nested within classrooms or repeated measures nested within individuals. These types of multilevel data violate the assumption of independence required to use traditional multiple linear regression methods. Recently, emerging research has begun to examine mediation in multilevel data using multilevel modeling (MLM; Bauer, Preacher, & Gil, 2006; Krull & MacKinnon, 1999, 2001; Pituch, Whittaker, & Stapleton, 2005; Zhang, Zyphur, & Preacher, 2009) and multilevel structural equation modeling (MSEM; Preacher, Zyphur, & Zhang, 2010; Preacher, Zhang, & Zyphur, 2011). In this paper, we discuss the strengths and limitations of each method, with a focus on the indirect effects when the predictor, mediator, and dependent variable are all collected at the lowest level of data. Based on a simulation study and empirical example, we also provide recommendations for applied researchers working on mediation in their own multilevel data.

Before discussing each method in more detail, it is important to establish appropriate terminology for these complex mediation designs. Krull and MacKinnon (2001) suggest a "Predictor-Mediator-Outcome" format, with a number indicating the level of data where each variable is located. Thus, for multilevel design with two levels (e.g., students nested within classrooms), "2-1-1" would mean the predictor is at level-2 (classroom) and the mediator and outcome are at level-1 (student), while "1-1-1" would mean all three variables are measured at level-1. In Figure 1 (adapted from Preacher et al. 2011), a 1-1-1 design for multilevel mediation is illustrated and associations among the three observed level-1 variables are separated into between-cluster and within-cluster effects. X, M, and Y represent the predictor, the mediator, and the outcome variable, respectively. Subscripts show if the variable represents the value for participant i in cluster j or average values for cluster j.

CONTACT Christina K. Zigler cek53@pitt.edu University of Pittsburgh, Physical Medicine and Rehabilitation, School of Medicine, Pittsburgh, 15213 USA; University of Pittsburgh, School of Education, 5924 Wesley W Posvar Hall, 230 South Bouquet Street, Pittsburgh, 15260 USA.
*Present address: RAND Corporation, Pittsburgh, PA, USA.

In terms of the estimated pathways, *a* refers to the coefficients of the paths from *X* to *M*, *b* to the coefficients of the paths from *M* to *Y* (after controlling for *X*), and *c*' to the coefficients of the paths from *X* to *Y* (after controlling for *M*). Subscripts of *b* and *w* indicate if the pathway is specified as between- or within-cluster, respectively.

For illustration purposes, we will use an empirical example throughout this manuscript. Assume that researchers are interested in learning if student engagement mediates the relationship between classroom climate and math achievement. Data were collected and all of the students within the sample were nested within specific classrooms, making the data structure multilevel. However, the researchers collected information from individual students, making a 1-1-1 design appropriate. In this example, the observed variables X_{ij} would refer to "student perception of classroom climate," M_{ij} to "student engagement in math learning," and Y_{ij} to "student math achievement" (for student *i* in classroom *j*). As in single-level mediation, the indirect pathways are likely of most interest to applied researchers and through MLM, researchers could estimate both between and within indirect effects. In a broad sense, if the between indirect effect was significant and positive, it would indicate that classrooms with more positive climates on average tend to have higher engagement at the classroom level, and in turn, this engagement would mediate the effect of classroom-average climate on collective achievement. If the within indirect was also significant and positive, it would indicate that overall, students within a classroom who rate the climate as more positive than average, tend to also report higher engagement than average, and in turn, that engagement would mediate the effects of self-reported classroom climate on their individual achievement. In addition to the indirect effects, it is possible to interpret the pathway estimates of *a*, *b*, and *c*' (Figure 1; described in more detail in the subsequent sections).

Mediation with MLMs

Krull and MacKinnon (1999) first introduced how to examine the multilevel mediation effects with a system of MLMs, expanding from the Baron–Kenny method developed for single-level data (Baron & Kenny, 1986). Later, researchers further adapted this approach to 1-1-1 mediation with random intercepts and slopes (Bauer et al., 2006; Pituch, Whittaker, & Stapleton, 2005), and other designs like 2-1-1 and 2-2-1 mediation (Krull & MacKinnon, 2001; Pituch, Stapleton, & Kang, 2006). MLMs include different error terms for each level of nesting and allow intercepts and slopes of level-1 variables to vary across clusters (Raudenbush & Bryk, 2002). Like in the Baron–Kenny approach, for mediation to be evaluated, at least two equations need to be estimated. The first equation models the association between the

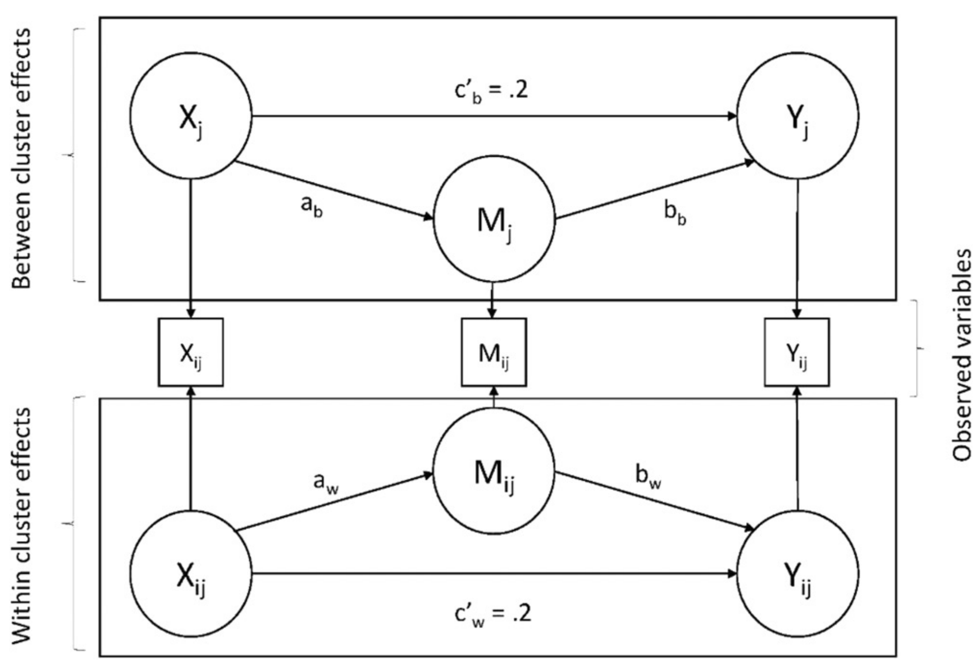

Figure 1. Path diagram depicting multilevel mediation in a 1-1-1 design where *X* represents the predictor, *M* the mediator, and *Y* the outcome.

predictor (X) with the mediator (M) through the main pathway of interest, a.

$$M_{ij} = \gamma_{M00} + aX_{ij} + u_{M0j} + u_{aj}X_{ij} + e_{Mij} \quad [1]$$

where M_{ij} and X_{ij} are the observation value for the level-1 mediator and predictor from participant i in cluster j. In this model, and the discussion of all subsequent models, all pathways are treated as random, meaning that the cluster-level error terms (for Equation 1; u_{M0j} and u_{aj}) allow the cluster intercept and pathway (in this case, a) to vary among clusters (level-2).

The second equation models the association between the mediator (M) and the predictor (X) with the outcome (Y) through pathways b and c'.

$$Y_{ij} = \gamma_{Y00} + bM_{ij} + c'X_{ij} + u_{Y0j} + u_{bj}M_{ij} + u_{cj}X_{ij} + e_{Yij} \quad [2]$$

where Y_{ij} is the observation from participant i in cluster j, and cluster level error terms, u_{Y0j}, u_{bj}, and $u_{cj,}$ again allow the cluster intercept and pathways (a, b, c') to vary across clusters. The average indirect effect of X on Y through M is calculated as the product of a and b when the pathways are not random (i.e., u_{Y0j}, u_{bj}, u_{cj} all have zero values), and as the product plus covariance of u_{aj} and u_{bj} when pathways are random (Bauer et al., 2006; see Table 1 for the equation for calculating the indirect effect for the CMM model with random pathways).

Before discussing about the subsequent models, it is important to note that MLM is not the only way to approach the analysis of multilevel data. Aggregation is an alternative method that was more commonly used before the advantages of MLM were well known. To aggregate data, individual level values are typically averaged for each cluster and then modeled using a traditional method. There are a number of drawbacks to using aggregation, the major ones being loss of information and decreased sample size (as n is now limited by the number of clusters). Even if researchers are most interested in the effects at level-2, aggregation can still result in biased estimates (Croon & van Veldhoven, 2007). For example, in a recent study looking at a contextual outcome, researchers found that aggregating individual data at level-2 would have led them to incorrectly reject their hypothesis (Becker, Breustedt, & Zuber, 2017). Specifically, for 1-1-1 designs, multilevel modeling for mediation has been found to be at greatest advantage over aggregation when intra-class correlations (ICCs) are high (for the mediator and outcome) and larger groups are included in analysis (Krull & MacKinnon, 2001). However, MLM is still limited in that it only allows estimation of one pathway (a instead of a_w and a_b), and thus does not provide separate estimates of the between and within indirect effects. Due to this limitation, when discussing mediation, it has been termed the "Conflated Multilevel Model" (CMM; Preacher et al., 2010, 2011). However, it is still a method that can be useful to model general mediation effects in multilevel data, and is often an applied researcher's first exposure to multilevel modeling.

Unconflated multilevel modeling (UMM), a strategy termed by Preacher et al. (2011), has been proposed as a straightforward modification to CMM (Krull & MacKinnon, 2001; MacKinnon, 2008; Zhang et al., 2009). Observed cluster means are used as a proxy of true cluster means in the population, which allows for separate estimations of between- and within-cluster effects (Zhang et al., 2009). Equation 1 now expands to;

$$\begin{aligned} M_{ij} = \gamma_{M00} + a_b X_{.j} + a_w(X_{ij} - X_{.j}) + u_{Moj} \\ + u_{a_wj}(X_{ij} - X_{.j}) + e_{Mij} \end{aligned} \quad [3]$$

where $X_{.j}$ is the cluster mean of X_{ij}, and pathway a is now parceled into the between- and within- cluster components; a_b and a_w. Similarly, Equation 2 is now expanded to be

$$\begin{aligned} Y_{ij} = \gamma_{Y00} + b_b M_{.j} + c'_b X_{.j} + b_w(M_{ij} - M_{.j}) \\ + c'_w(X_{ij} - X_{.j}) + u_{Yoj} + u_{b_wj}M_{ij} + u_{c_wj}X_{ij} + e_{Yij} \end{aligned} \quad [4]$$

where $M_{.j}$ is the cluster mean of M_{ij}, and pathways b and c' are parceled into between (b_b, c'_b) and within cluster components (b_w, c'_w). Because the within cluster effects are considered random, calculation for the within indirect effect incorporates the covariance of u_{a_wj} and u_{b_wj} (Table 1; Bauer et al., 2006; Kenny et al., 2003). This means that if the within pathways are

Table 1. Calculation of indirect effects for three different multilevel mediation modeling methods.

Method	Type of indirect effect	Calculation of indirect effect	Corresponding model equations in text
CMM	"Conflated"	$\text{ind}_{CMM} = a^*b +$ cov (u_{aj}, u_{bj})	1 & 2
UMM	Within	$\text{ind}_{UMM,w} = a_w{}^*b_w +$ cov (u_{a_wj}, u_{b_wj})	3 & 4
	Between	$\text{ind}_{UMM,b} = a_b * b_b$	
MSEM	Within	$\text{ind}_{MSEM,,w} = \mu_{B_{MX_j}} * \mu_{B_{YM_j}} + \text{cov}(\beta_{MX_j}, \beta_{YM_j})$	5–7
	Between	$\text{ind}_{MSEM,b} = \beta_{MX} * \beta_{YM}$	

Note: ind = indirect effect, a = coefficient of pathway from predictor X to mediator M, b = coefficient of pathway from mediator M to outcome Y (after adjusting for X), subscript w = within pathway, subscript b = between pathway, cov = covariance, CMM = conflated multilevel model, UMM = unconflated multilevel model, MSEM = multilevel structural equation model. Note that when MSEM with random slopes is estimated in M*plus* 7, $\beta_{MX} = a_w + a_b$, and $\beta_{YM} = b_w + b_b$ where estimated a_b and b_b are compositional effects; that is, the prediction from cluster level variables after controlling for within cluster effects (Raudenbush & Bryk,2002).

both 0, the calculated within indirect effect can still be nonzero if there is nonzero covariance of $u_{a_{wj}}$ and $u_{b_{wj}}$. The between-cluster indirect effect is simply the product of a_b and b_b. It is important to note that the within pathways interpretation has changed somewhat from CMM, as the variables now reflect the deviation from cluster averages (e.g., M_{ij}–$M_{.j}$). Despite the improvement over CMM in separating between- and within-cluster indirect effects, using observed cluster means as the proxy of true cluster mean can result in a bias of between indirect effects toward the within effect to the extent that ICCs are low and cluster sizes are small (Preacher et al., 2011).

Multilevel mediation using multilevel structural equation modeling

Preacher et al. (2010) proposed using a general MSEM approach (Muthén & Asparouhov, 2008) for multilevel mediation, which has many stated advantages over CMM/UMM. It incorporates multilevel mediation in a general framework via a multivariate path diagram. MSEM treats the cluster level component of the level-1 variables as latent and can test between- and within-cluster pathways simultaneously (Ludtke et al., 2008; Preacher et al. 2010, 2011). A MSEM model for multilevel 1-1-1 mediation with random slopes, as reported in Preacher et al. (2010), includes a measurement model, a within structural model and a between structural model, as shown in Equations 5–7.

Measurement model $\boldsymbol{Y}_{ij} = \boldsymbol{\Lambda}\boldsymbol{\eta}_{ij}$

$$= \begin{bmatrix} X_{ij} \\ M_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{X_{ij}} \\ \eta_{M_{ij}} \\ \eta_{Y_{ij}} \\ \eta_{X_j} \\ \eta_{M_j} \\ \eta_{Y_j} \end{bmatrix} \quad [5]$$

Within Structural Model $\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \boldsymbol{B}_j\boldsymbol{\eta}_{ij} + \boldsymbol{\xi}_{ij}$

$$= \begin{bmatrix} \eta_{X_{ij}} \\ \eta_{M_{ij}} \\ \eta_{Y_{ij}} \\ \eta_{X_j} \\ \eta_{M_j} \\ \eta_{Y_j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \alpha_{\eta_{X_j}} \\ \alpha_{\eta_{M_j}} \\ \alpha_{\eta_{Y_j}} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ B_{MX_j} & 0 & 0 & 0 & 0 & 0 \\ B_{YX_j} & B_{YM_j} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{X_{ij}} \\ \eta_{M_{ij}} \\ \eta_{Y_{ij}} \\ \eta_{X_j} \\ \eta_{M_j} \\ \eta_{Y_j} \end{bmatrix} + \begin{bmatrix} \xi_{X_{ij}} \\ \xi_{M_{ij}} \\ \xi_{Y_{ij}} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad [6]$$

Between Structural Model

$$\boldsymbol{\eta_j} = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta_j} + \boldsymbol{\zeta_j}$$

$$= \begin{bmatrix} \alpha_{\eta_{X_j}} \\ \alpha_{\eta_{M_j}} \\ \alpha_{\eta_{Y_j}} \\ B_{MX_j} \\ B_{YM_j} \\ B_{YX_j} \end{bmatrix} = \begin{bmatrix} \mu_{\alpha_{\eta_{X_j}}} \\ \mu_{\alpha_{\eta_{M_j}}} \\ \mu_{\alpha_{\eta_{Y_j}}} \\ \mu_{B_{MX_j}} \\ \mu_{B_{YM_j}} \\ \mu_{B_{YX_j}} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{MX} & 0 & 0 & 0 & 0 & 0 \\ \beta_{YX} & \beta_{YM} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{\eta_{X_j}} \\ \alpha_{\eta_{M_j}} \\ \alpha_{\eta_{Y_j}} \\ B_{MX_j} \\ B_{YM_j} \\ B_{YX_j} \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha_{\eta X_j}} \\ \zeta_{\alpha_{\eta M_j}} \\ \zeta_{\alpha_{\eta Y_j}} \\ \zeta_{B_{MX_j}} \\ \zeta_{B_{YM_j}} \\ \zeta_{B_{YX_j}} \end{bmatrix} \quad [7]$$

The measurement model partitions each observed variable into within and between components that are treated as a latent variable. This is similar to the variance partition that occurs within UMM, but is notationally represented as a measurement model in MSEM. In the within structural model, the $\boldsymbol{\alpha}_j$ vector represents estimates of the true cluster means, similar to the advantage of including $M_{.j}$ in Equation 4. $\boldsymbol{B}_j$ is a matrix of structural regression parameters including the random within pathways of a, b, and c', represented by nonzero entries of B_{MX_j}, B_{YM_j}, B_{YX_j}, and $\boldsymbol{\xi}_{ij}$ is a vector of latent within-cluster residuals. In the between structural model, the $\boldsymbol{\eta_j}$ vector contains the random effects that potentially vary at the cluster level and vector $\boldsymbol{\mu}$ contains the population means of the random within-cluster intercepts and slopes over all clusters. Nonzero entries in the $\boldsymbol{\beta}$ matrix, $\beta_{MX}, \beta_{YM}, \beta_{YX}$ are the coefficients for the between pathways of a, b, and c'. The within-cluster residual vector, $\boldsymbol{\xi_{ij}}$, and the between-cluster residual vector, $\boldsymbol{\zeta_j}$, are both multivariate normal with a mean of 0 and covariance matrix $\boldsymbol{\Psi}_{\xi}$ and $\boldsymbol{\Psi}_{\zeta}$ respectively (Preacher et al., 2010). The within indirect effect is calculated as $\mu_{B_{MX_j}} {}^{*} \mu_{B_{YM_j}} +$ cov(B_{MX_j},B_{YM_j}), which is essentially the same as the equation for UMM, and the between indirect effect as $\beta_{MX}{}^{*}\beta_{YM}$(Table 1). The existing CMM and UMM approaches may be expressed as special cases of MSEM by constraining $\boldsymbol{B} = \boldsymbol{\beta}$, or replacing η_{X_j}, η_{M_j}, η_{Y_j} with observed cluster means and $\eta_{X_{ij}}$, $\eta_{M_{ij}}$, $\eta_{Y_{ij}}$ with cluster mean centered variables. Please note for ease in understanding, for the rest of the document, we will use a, b, and c' as defined above to refer to the corresponding pathways for all three modeling methods, even though the reader should note that they represent different parameters from each method.

The MSEM model has been shown to perform better than CMM and UMM in regards to efficiency, bias, and confidence interval (CI) coverage in a 2-1-1 design with both fixed and random *b* slopes under certain data conditions (when sample size, cluster size, and ICCs were large; Preacher et al., 2011). Adequate power for detecting significant between indirect effects was also reached with large enough sample sizes (Preacher et al., 2011).

However, in practice, the sample size requirements for MSEM can be problematic to meet. In a random sampling of empirical studies citing Preacher et al. (2010), more than half of papers (56%) used <50 clusters (McNeish, 2017). In an accompanying simulation study, McNeish (2017) found that MSEM had poor CI coverage (<92%) when fewer than 50 clusters were used, while the multilevel model showed no such issues (coverage was between acceptable range as per Bradley (1978) for all conditions). Therefore, despite the attractiveness of MSEM and its popularity as a technique (McNeish, 2017), there are certain scenarios when one modeling method would be more appropriate than the others. In regards to 1-1-1 mediation, there are significant gaps in the published literature that prevent applied researchers from making educated decisions regarding which model to employ in their own specific research design. Established comparisons between modeling methods published by Preacher et al. (2011) and McNeish (2017) focused on simulating data from 2-1-1 designs.

In the following sections, we first conduct a Monte Carlo simulation study to compare UMM, CMM, and MSEM for estimating multilevel mediation in various combinations of sample sizes, ICC's, strengths of indirect effects, and the existence (or not) of between indirect effects. Then, we illustrate potential differences among the three methods using an empirical example. Within this example, we also provide a deeper conceptual explanation of the indirect effects and differences in interpretation between methods.

Simulation

Simulation design

A Monte Carlo study was performed to compare the three multilevel mediation modeling methods (CMM, UMM, and MSEM) on CI coverage, power, and relative bias under varying data conditions. We targeted a 1-1-1 design with random slopes for our simulation study. Although theoretically simple, a 1-1-1 design allows for the exploration of both the between and within components of the indirect effects and pathways (Zhang et al., 2009). It also is very common in applied studies. In a recent random sample of peer-reviewed studies citing Preacher et al. (2010) and using MSEM for multilevel mediation, 41% tested at least one 1-1-1 design (McNeish, 2017). Also, in contrast to designs with the mediator or outcome at level two (i.e., 1-2-1 or 1-2-2), researchers modeling with this type of data structure are able to explicitly choose between MLM (either CMM or UMM) and MSEM. MSEM has also not yet been evaluated against CMM and UMM when estimating a 1-1-1 design. To the authors' knowledge, so far only Bauer et al. (2006) included random slopes of all three level-1 mediation pathways (*a*, *b* and *c*'), but they adopted CMM and thus, did not estimate separate within and between indirect effects. In addition, this detailed exploration of all random slopes could also not be evaluated in a 2-1-1 design (as seen in Appendix of Preacher et al., 2011), due to the predictor being only measured at level-2. The current study intends to provide a more comprehensive comparison of different models for multilevel mediation while the within-cluster mediation pathways are random.

CMM is known to have a number of limitations when modeling mediation relationships in multilevel data. However, it was important to include in this simulation study for a number of reasons. First, it is a very straightforward model that is commonly used in the applied literature and is often an applied researcher's first exposure to multilevel modeling. Second, we wanted to explore contexts in which CMM (or UMM) might be preferred over MSEM, especially when clustering effect is only a nuisance effect and a between indirect effect may not exist. In this situation, an applied researcher might practically prefer a CMM model over UMM or MSEM, as it does not explicitly estimate between pathways.

As in Preacher et al's study (2011), data were generated using an MSEM framework (Equations 5–7) as it can explicitly simulate between- and within-cluster pathways. All variables were continuous and errors were generated from normal distributions. The variances and covariances of the within and between error terms were adapted from values used in Bauer et al. (2006). Specifically, the within structural model errors were represented as,

$$Cov\begin{pmatrix} \xi_{X_{ij}} \\ \xi_{M_{ij}} \\ \xi_{Y_{ij}} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & .65 & 0 \\ 0 & 0 & .45 \end{bmatrix}$$

Table 2. Varied factors when simulating 1-1-1 multilevel mediation.

Estimation method	n_2	n_1	Cluster effect for *X*, *M*, and *Y*	Strength of between indirect effect	Within Cov_{ab}
CMM	30	20	$\rho_{all} = .1$	Small: 0.01	−0.113
UMM	60	40	$\rho_{all} = .3$	Medium: 0.09	0
MSEM	100			Large: 0.36	0.113
	500				

Note: CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; n_2 = number of clusters; n_1 = cluster size; ρ_{all} = ICC = interclass correlation coefficient; Cov_{ab} = covariance of within cluster *a* and *b* pathways.

while the residual covariance structure of the random effects was,

$$\text{Cov}\begin{pmatrix} \zeta_{\alpha_{\eta X_j}} \\ \zeta_{\alpha_{\eta M_j}} \\ \zeta_{\alpha_{\eta Y_j}} \\ \zeta_{B_{MX_j}} \\ \zeta_{B_{YM_j}} \\ \zeta_{B_{YX_j}} \end{pmatrix} = \begin{bmatrix} \tau_X & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_M & 0 & 0 & 0 & 0 \\ 0 & 0 & \tau_Y & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.16 & \sigma_{ab} & 0 \\ 0 & 0 & 0 & \sigma_{ab} & 0.16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.04 \end{bmatrix}$$

The covariance matrix of cluster level error terms suggests that the variances of the random within-cluster pathways of *a*, *b*, and *c*' were 0.16, 0.16, and 0.04 respectively, with the covariance between the two pathways *a* and *b* taking on varied values (described more in subsequent sections). Also, the cluster level intercept variances of *X*, *M*, and *Y* (i.e., τ_X, τ_M, τ_Y) were varied depending on ICC values.

We simulated data under three mediation conditions. For the first mediation condition, we simulated data with mediation effects present at between- and within-cluster levels. Thus, both between and within pathways were explicitly generated to be nonzero. The second mediation condition simulated only the within-cluster pathways, with the between pathways set to 0 (i.e., $\boldsymbol{\beta}=\mathbf{0}$ in Equation 7), to mimic research scenarios where the only interest is in within-cluster level mediation, and clustering effects are considered to be confounding. No mediation was simulated in the third condition so that we could explore Type I Error. All three are described in more detail in the following sections.

Condition 1: True multilevel mediation

In the true multilevel mediation condition, there were six manipulated factors (Table 2). The included levels of ICC for *X*, *M*, and *Y* reflect small and large clustering effects (ICC = .1 and .3) commonly encountered in educational studies (Hedges & Hedberg, 2007; Hox & Maas, 2001) and encompass the general range simulated in Preacher et al. (2011). The level-1 residual variances of *X*, *M*, and *Y* were fixed to be 1, .65, and .45. As ICC is defined as the cluster-level intercept variances of *X*, *M*, and *Y* (i.e., τ_X, τ_M, τ_Y) over the total variance for each variable (i.e., the sum of level-1 residual variance and the cluster level intercept variance), the cluster level intercept variances of *X*, *M*, and *Y* were calculated as ICC/(1-ICC), .65ICC/(1-ICC), and .45ICC/(1-ICC), respectively.

As in previous studies, we also varied the number of clusters and cluster size separately (as per Preacher et al., 2011) and used a wide range of sample sizes commonly found in applied studies, including smaller sample sizes (as per McNeish, 2017). The within-cluster pathways a_w and b_w were set to .3, as we explicitly wanted to focus on model performance when the between indirect effect varied in contrast to the within indirect effect. However, the covariance of the a_w and b_w pathways was varied, affecting the "true" value of the within indirect effect. The true effect would be −.023, .09, and .203 when the covariance of a_w and b_w was −.113, 0, and .113. We also varied the strength of the between indirect pathways at .1, .3, and .6, reflecting small, moderate, and large indirect effects, following Fritz and MacKinnon (2007). Such variation in between indirect effects represent the situations when the between indirect effect is smaller, equal, and larger than the within indirect effect. For this mediation condition, the c' pathways (both between and within) were set to be .2, simulating partial mediation, which is more likely to be found in practice than complete mediation.

Condition 2: Within-cluster mediation only

For the second mediation condition, we set the between pathways, a_b, b_b, and c'_b, to be 0. As in the first condition, the within cluster pathways were set to $a_w = b_w = .3$, and $c'_w = .2$. For this condition, however, we further explored the effect of ICC. As in condition 1, we included small and large ICCs for *X*, *M*, and *Y*, but also included small and large ICCs for only *Y*, setting ICCs for *X* and *M* to be 0. This mimics research scenarios when clustering effects only exists for the outcome variable, but not for the predictor and mediator. For this condition, we also explored the results over the levels of sample size and covariance of a_w and b_w, as shown in Table 2.

Condition 3: No mediation condition

A final condition was included as a way to evaluate Type I Error of the modeling methods. In this condition, we set all between and within pathways (a, b, c') and covariance of a_w and b_w to zero. We also included the same four ICC conditions, as seen in condition 2 (small and large for X, M, and Y; small and large for only Y). For this condition, we report the resulting Type I Error for each modeling method over the different levels of sample size at level-1 and level-2 (Table 2).

Outcome measures

For the first simulated mediation condition, three outcome measures were of interest, including CI coverage, relative bias, and power, for both between and within indirect effects estimated by CMM, UMM, and MSEM. For the second mediation condition (only within-cluster mediation was simulated), the same three outcome measures were of interest for the within indirect effect. For the between indirect effect, we assessed only Type I Error. For the third and final mediation condition (all pathways were 0), we assessed Type I Error for both the between and within effects.

Computation of confidence intervals. We computed a symmetric 95% CI for both between or within indirect effects using Equation 8 (Preacher & Selig, 2010).

$$95\%\ \mathrm{CI}\{\mathrm{upper},\ \mathrm{lower}\} = \widehat{\mathrm{ind}} \pm 1.96 * \mathrm{SE}\left(\widehat{\mathrm{ind}}\right) \qquad [8]$$

where $\widehat{\mathrm{ind}}$ represents the estimated indirect effect (either between or within), and SE is the corresponding standard error, as calculated by M*plus* 7. The standard error of the within indirect effect with the presence of random slopes is calculated following Equation 9 in Bauer et al. (2006, p. 147).

The method used to calculate the CI assumes that the indirect effect is normally distributed, and thus symmetric about the point estimate. Other methods have been recommended to calculate CI that do not assume the normality of the indirect effect, including the product method (MacKinnon, Fritz, Williams, & Lockwood, 2007; Tofighi & MacKinnon, 2011) and the Monte Carlo method (Preacher & Selig, 2012). We chose to adopt the current method for this study, as this kind of CI suffices for comparing methods in a simulation (Preacher et al., 2011), and was found not to differ significantly from the product method and Monte Carlo method in terms of CI coverage rates while assessing multilevel mediation when both level-1 and level-2 sample sizes were small ($n_1 < 15$, $n_2 = 25$; McNeish, 2017).

Type 1 Error rate. If the 95% CI contained zero, the effect was considered "nonsignificant." Rates of incorrectly identified significant effects were calculated over all replications for the simulation conditions. For the "no mediation" condition, we reported Type I Error rates for both the between and within effects by sample size, ICC, and modeling method. For the second mediation condition, where only within pathways were simulated, we reported Type I Error rates for the between indirect effects by sample size, covariance, ICC, and modeling method.

CI coverage rate. The 95% CI coverage rate was calculated as the percentage of replications where the population value was located within the estimated CI. For a 95% CI, it is ideal to have a coverage rate of .95; Bradley (1978) suggested that CI coverage rates <92.5% or >97.5% are problematic. However, it is important to note that prior research has found that when assuming a symmetric CI with multilevel modeling, coverage of the indirect effect tends to be <95% even under the best of circumstances, and could decrease to as low as .6 when sample size and indirect effect size are small (McNeish, 2017; Preacher et al., 2011).

Power. Power was calculated as the percentage of replications in which the estimated nonzero indirect effect was significant, i.e., the 95% CI of the indirect effect did not contain zero. Power rates were reported for nonzero between and within indirect effects for all simulation conditions.

Relative bias. Relative bias of the respective indirect effect was calculated for each nonzero condition using the following equation:

$$\mathrm{bias} = \frac{(\widehat{\mathrm{ind}} - \mathrm{true})}{\mathrm{true}} \qquad [9]$$

where "true" indicates the true value based on the simulation study design. Calculating relative bias in this way provides a conceptualization of the percentage of bias out of the total value of the "true" parameter. Positive values mean that the modeling method over-estimated the parameter, while negative values mean the parameter was under-estimated.

Relative bias in standard error estimation. For all simulation conditions, the accuracy of standard error estimation was explored for each modeling method. Examination of standard error bias was especially important due to our utilization of the delta method, which uses the estimated standard errors to construct CIs and determine statistical significance. Using the empirical standard deviation of the relevant indirect effect estimates (between or within) for each simulation condition as the "true" standard error, we defined "relative bias" for the standard errors as the ratio of difference between average standard error estimate and the true standard error value over the true standard error. Positive relative bias values indicated that the standard errors were overestimated, while negative ones indicated underestimation.

Software details

The entire simulation was run in SAS 9.4 (SAS Institute, 2012). One thousand data sets were generated from a 1-1-1 mediation design with random slopes under various simulation conditions. An automization program was written in SAS that utilized an X command to call M*plus* 7.3 (Muthén & Muthén, 1998–2017). The M*plus* syntax used to estimate the 1-1-1 design using CMM, UMM, and MSEM was adopted from the supplemental materials provided for Preacher et al.'s 2010 and 2011 papers.

Simulation results

Results for condition 1: True multilevel mediation

CI Coverage. Confidence interval coverage of *within* indirect effects were in the range for acceptable coverage (92.5–97.5%; Bradley, 1978) for most conditions under all three modeling methods (Table 3). Coverage rates were at their lowest (<.9) when level-2 sample size was 30 and the covariance was positive. For *between* indirect effects, rates of CI coverage were overall higher for MSEM when ICCs were .1, and they were comparable when ICCs were .3 (Table 4). However, coverage was problematic when the between indirect effect was small ($\text{ind}_b = .01$), as the rates were higher than the range generally considered to be acceptable (>97.5%). Rates for between indirect effects were higher for this condition when using UMM as well, although the value did not exceed the acceptable range in all conditions. When the between indirect pathways were of medium strength, MSEM showed lower than ideal coverage when the covariance of *a* and *b* was negative. When between indirect effects were large ($\text{ind}_b = .36$), CI coverage was not within the acceptable range for any of the samples size conditions for UMM, but MSEM had much better results (although coverage rates were best when the covariance was positive).

Bias. The relative bias for within indirect effects was generally acceptable (<5%) for CMM, UMM, and MSEM over most of the simulation conditions

Table 3. Confidence interval coverage for **within** indirect effects in 1-1-1 designs for CMM, UMM, and MSEM (where $a_w = .3$ and $b_w = .3$).

			CMM									UMM									MSEM								
			$\text{ind}_b = .36$			$\text{ind}_b = .09$			$\text{ind}_b = .01$			$\text{ind}_b = .36$			$\text{ind}_b = .09$			$\text{ind}_b = .01$			$\text{ind}_b = .36$			$\text{ind}_b = .09$			$\text{ind}_b = .01$		
cov_{ab}	n_2	n_1	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
$\rho = .1$	30	20	.90	0.91	0.82	0.89	0.92	0.84	0.91	0.93	0.81	.92	0.92	0.84	0.90	0.92	0.85	0.92	0.93	0.83	.91	0.91	0.81	0.89	0.92	0.84	0.91	0.93	0.82
	60	20	0.92	0.95	0.92	0.93	0.95	0.91	0.93	0.94	0.91	0.94	0.95	0.93	0.94	0.94	0.93	0.94	0.94	0.92	0.92	0.95	0.91	0.93	0.94	0.91	0.93	0.94	0.91
	100	20	0.92	0.94	0.94	0.92	0.94	0.91	0.94	0.94	0.91	0.95	0.93	0.94	0.95	0.94	0.93	0.95	0.95	0.93	0.93	0.94	0.93	0.93	0.94	0.91	0.94	0.94	0.91
	500	20	0.86	0.94	0.92	0.89	0.96	0.92	0.92	0.94	0.89	0.94	0.95	0.94	0.95	0.95	0.95	0.95	0.94	0.95	0.89	0.95	0.91	0.89	0.96	0.92	0.90	0.94	0.91
	30	40	0.90	0.93	0.84	0.90	0.92	0.87	0.93	0.93	0.86	0.90	0.93	0.86	0.91	0.92	0.88	0.94	0.93	0.87	0.90	0.93	0.84	0.90	0.91	0.86	0.93	0.93	0.86
	60	40	0.92	0.94	0.92	0.93	0.95	0.92	0.93	0.94	0.91	0.94	0.93	0.92	0.94	0.95	0.92	0.94	0.95	0.93	0.93	0.94	0.91	0.93	0.95	0.92	0.93	0.95	0.92
	100	40	0.93	0.94	0.93	0.93	0.95	0.93	0.93	0.95	0.93	0.94	0.94	0.94	0.94	0.95	0.94	0.93	0.95	0.93	0.93	0.94	0.93	0.93	0.95	0.93	0.92	0.95	0.93
	500	40	0.90	0.95	0.94	0.92	0.95	0.94	0.94	0.95	0.93	0.95	0.95	0.96	0.94	0.94	0.96	0.95	0.95	0.95	0.92	0.95	0.93	0.92	0.94	0.94	0.94	0.96	0.93
$\rho = .3$	30	20	0.90	0.92	0.83	0.91	0.92	0.82	0.91	0.91	0.80	0.92	0.91	0.83	0.92	0.92	0.86	0.93	0.91	0.83	0.90	0.91	0.82	0.91	0.92	0.82	0.91	0.91	0.81
	60	20	0.91	0.95	0.91	0.92	0.95	0.92	0.92	0.94	0.90	0.94	0.94	0.92	0.94	0.95	0.94	0.94	0.94	0.93	0.92	0.95	0.91	0.92	0.95	0.92	0.92	0.94	0.91
	100	20	0.90	0.95	0.92	0.93	0.95	0.91	0.94	0.94	0.91	0.94	0.95	0.94	0.95	0.95	0.93	0.95	0.93	0.94	0.91	0.94	0.91	0.93	0.94	0.91	0.93	0.94	0.92
	500	20	0.81	0.94	0.92	0.87	0.95	0.90	0.91	0.93	0.88	0.95	0.94	0.95	0.95	0.95	0.96	0.95	0.94	0.94	0.86	0.94	0.90	0.87	0.95	0.90	0.88	0.94	0.90
	30	40	0.90	0.93	0.85	0.92	0.93	0.83	0.92	0.94	0.86	0.91	0.92	0.86	0.92	0.92	0.85	0.92	0.93	0.88	0.90	0.93	0.85	0.92	0.92	0.84	0.92	0.94	0.86
	60	40	0.93	0.95	0.93	0.93	0.94	0.92	0.93	0.94	0.91	0.95	0.95	0.93	0.94	0.94	0.93	0.93	0.94	0.93	0.94	0.95	0.93	0.93	0.94	0.92	0.92	0.95	0.92
	100	40	0.92	0.95	0.93	0.93	0.93	0.93	0.94	0.95	0.92	0.94	0.95	0.94	0.95	0.93	0.94	0.94	0.95	0.93	0.92	0.95	0.92	0.93	0.93	0.93	0.93	0.95	0.93
	500	40	0.90	0.95	0.94	0.91	0.95	0.92	0.94	0.96	0.92	0.95	0.95	0.95	0.94	0.94	0.94	0.95	0.96	0.95	0.91	0.95	0.93	0.91	0.95	0.92	0.93	0.96	0.93

Note: ind_b = strength of between pathways; n_2 = number of groups; n_1 = group size; ρ = population interclass correlation coefficient; CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_{ab} = covariance of *a* and *b* pathways where 1 = –.113, 2 = 0, and 3 = .113.

Table 4. Confidence interval coverage for between indirect effects in 1-1-1 designs for UMM and MSEM.

				UMM									MSEM								
				$\text{ind}_b = .36$			$\text{ind}_b = .09$			$\text{ind}_b = .01$			$\text{ind}_b = .36$			$\text{ind}_b = .09$			$\text{ind}_b = .01$		
ρ	n_2	n_1	cov_{ab}	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113
.1	30	20		.61	.55	.54	.87	.86	.87	.98	.98	.98	.84	.86	.88	.91	.91	.91	.99	1.00	1.00
	60	20		.65	.66	.67	.90	.90	.91	.96	.97	.96	.86	.91	.93	.88	.90	.93	.99	.99	.99
	100	20		.54	.54	.57	.90	.91	.93	.96	.96	.95	.85	.90	.94	.87	.90	.93	.98	.98	.98
	500	20		.06	.05	.08	.92	.94	.94	.68	.56	.47	.66	.92	.94	.87	.95	.95	.91	.93	.94
	30	40		.65	.66	.66	.86	.87	.88	.98	.98	.98	.72	.80	.88	.86	.89	.89	.99	.99	.99
	60	40		.78	.80	.80	.90	.92	.92	.96	.98	.97	.81	.89	.94	.87	.90	.91	.98	.98	.98
	100	40		.78	.75	.78	.91	.93	.92	.95	.96	.96	.81	.89	.94	.88	.92	.94	.96	.97	.97
	500	40		.36	.41	.42	.93	.94	.96	.90	.84	.79	.50	.87	.94	.85	.93	.95	.94	.95	.95
.3	30	20		.77	.74	.73	.88	.88	.87	.98	.98	.98	.75	.83	.86	.86	.89	.89	.99	.98	.99
	60	20		.87	.86	.86	.90	.91	.91	.96	.97	.96	.84	.90	.94	.87	.90	.93	.98	.98	.98
	100	20		.86	.87	.87	.91	.93	.92	.95	.95	.96	.83	.92	.94	.88	.92	.92	.95	.97	.95
	500	20		.73	.75	.73	.93	.94	.95	.94	.95	.93	.58	.91	.93	.87	.94	.94	.92	.94	.95
	30	40		.80	.80	.81	.87	.87	.86	.98	.98	.98	.71	.81	.86	.86	.86	.87	.98	.99	.98
	60	40		.90	.90	.89	.92	.92	.90	.96	.96	.96	.80	.90	.93	.87	.90	.91	.97	.97	.98
	100	40		.90	.90	.89	.93	.91	.92	.94	.94	.94	.77	.90	.93	.87	.92	.93	.96	.96	.96
	500	40		.88	.89	.88	.94	.94	.94	.95	.96	.95	.42	.87	.93	.85	.94	.95	.94	.96	.95

Note. ind_b = strength of between indirect effect; n_2 = number of groups; n_1 = group size; ρ = population interclass correlation coefficient; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_ab = covariance of a and b pathways.

(Table 5). However, it is important to note that for three methods, bias was very large when the covariance of a and b was negative. This problematic bias was made worse with CMM and MSEM, with lower sample sizes, and when the corresponding between-pathways were large. In some cases, bias reached over 50% of the true pathway. Negative covariance did not have nearly as strong an effect on bias when using UMM to estimate the within indirect effects; although it was still highest for those conditions. Relative bias only reached >10% when the level-2 sample size was small ($n_2 = 30$). For all methods, relative bias in within indirect effects seemed to be generally negative, so the parameters tended to be underestimated.

UMM and MSEM showed different patterns of relative bias for between indirect effects over the simulation factors. Relative bias for between indirect effects was overall high with many simulation conditions reaching >.05. Bias was the highest when using UMM if the cluster level effect, sample sizes, and between pathways were small (Table 6). Although bias was also higher in these conditions for MSEM (when compared to other conditions using MSEM), it did not reach the same magnitude; none of the cells surpassed 50% bias in the true parameter. Overall, UMM showed more acceptable levels of bias (generally <10%) when the between pathways were equivalent to the within pathways (a & b for both between and within pathways set to .3). When estimated with MSEM, relative bias in between indirect effects seemed to be more strongly affected by the covariance of a and b pathways. Under the 0 covariance conditions, the bias was lowest for all sample sizes and ICCs, except when the between indirect effect was small ($\text{ind}_b = .01$).

There also was an interesting relationship between the strength of the within and between pathways and the direction of relative bias of between effects when using UMM (Table 6). When the between pathways were stronger than within pathways (including when negative covariance decreased the within indirect effect estimate), bias in the between indirect effects was negative for all simulation conditions. However, when the between indirect effects were equivalent to or less than the within indirect effects, the relative bias for the between indirect effects was positive. Since UMM uses the group means of level-1 values as proxies for population cluster averages (which in turn are used to compute between indirect effects), it is not surprising that the estimates of the between indirect effects are being biased in the direction of the within pathways. Although not as extreme, this overall pattern also generally held for MSEM when the covariance of the a and b pathways was not negative.

Power. For all methods, power of the within indirect effects was directly related to the sample size and the strength of the true value of the within indirect effect, and this relationship was similar for CMM, UMM, and MSEM (Table 7). Power increased as the sample size (at both level 1 and 2) and the strength of the within indirect effect increased (as per covariance of a_w and b_w). For all methods, acceptable power ($\geq .8$) was reached when the covariance was positive for all other levels of simulation factors. When using any of the methods, acceptable power was reached for the 0 covariance condition (true within indirect effect = .09) when at least 60 groups were present. None of the methods reached acceptable power to detect within indirect effects when covariance was

Table 5. Relative bias for within indirect effects in 1-1-1 designs for CMM, UMM, and MSEM strategies (where $a_w = .3$ and $b_w = .3$).

			CMM									UMM									MSEM								
			$ind_b = .36$			$ind_b = .09$			$ind_b = .01$			$ind_b = .36$			$ind_b = .09$			$ind_b = .01$			$ind_b = .36$			$ind_b = .09$			$ind_b = .01$		
cov_{ab}	n_2	n_1	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
$\rho = .1$	30	20	−.75	−.08	−.19	−.56	−.06	−.19	−.57	−.09	−.20	−.42	−.11	−.17	−.31	−.06	−.17	−.40	−.07	−.17	−.68	−.10	−.20	−.56	−.06	−.19	−.64	−.07	−.20
	60	20	−.38	−.02	−.03	−.25	−.01	−.03	−.15	−.02	−.05	−.00	−.01	−.01	−.01	−.01	−.01	−.05	−.00	−.01	−.29	−.00	−.04	−.25	−.01	−.03	−.22	−.00	−.04
	100	20	−.41	−.03	−.02	−.29	−.00	−.03	−.14	−.01	−.03	−.03	−.00	−.00	−.02	−.00	−.00	−.06	−.00	−.00	−.33	−.01	−.03	−.29	−.00	−.03	−.21	−.01	−.02
	500	20	−.38	−.03	−.02	−.28	−.00	−.03	−.20	−.02	−.04	−.01	−.00	−.00	−.00	−.00	−.00	−.00	−.00	−.00	−.30	−.01	−.03	−.28	−.00	−.03	−.28	−.00	−.03
	30	40	−.30	−.03	−.14	−.36	−.05	−.15	−.28	−.07	−.14	−.12	−.05	−.13	−.22	−.05	−.14	−.17	−.06	−.13	−.27	−.04	−.14	−.36	−.05	−.15	−.32	−.06	−.14
	60	40	−.17	−.01	−.02	−.12	−.00	−.02	−.08	−.00	−.03	−.04	−.00	−.01	−.03	−.00	−.00	−.04	−.01	−.01	−.14	−.00	−.02	−.12	−.00	−.02	−.12	−.01	−.02
	100	40	−.22	−.01	−.01	−.16	−.01	−.01	−.13	−.01	−.03	−.00	−.00	−.00	−.00	−.01	−.01	−.01	−.00	−.01	−.18	−.00	−.02	−.16	−.01	−.01	−.17	−.01	−.02
	500	40	−.23	−.02	−.01	−.15	−.01	−.02	−.11	−.01	−.02	−.01	−.00	−.00	−.00	−.01	−.00	−.01	−.00	−.00	−.19	−.00	−.02	−.15	−.01	−.02	−.16	−.00	−.02
$\rho = .3$	30	20	−.74	−.07	−.19	−.65	−.08	−.19	−.59	−.10	−.21	−.31	−.10	−.17	−.36	−.08	−.16	−.38	−.08	−.17	−.66	−.09	−.20	−.66	−.08	−.20	−.68	−.08	−.20
	60	20	−.51	−.04	−.04	−.27	−.00	−.05	−.18	−.02	−.06	−.04	−.00	−.01	−.07	−.00	−.01	−.07	−.01	−.01	−.42	−.01	−.05	−.27	−.00	−.05	−.28	−.01	−.05
	100	20	−.51	−.04	−.02	−.29	−.00	−.03	−.27	−.01	−.05	−.03	−.00	−.00	−.04	−.00	−.00	−.02	−.01	−.00	−.42	−.01	−.04	−.29	−.00	−.03	−.37	−.01	−.04
	500	20	−.49	−.04	−.03	−.34	−.00	−.04	−.25	−.02	−.05	−.01	−.00	−.00	−.00	−.00	−.00	−.01	−.00	−.00	−.40	−.01	−.04	−.34	−.00	−.04	−.35	−.00	−.04
	30	40	−.49	−.04	−.14	−.32	−.04	−.16	−.25	−.03	−.15	−.27	−.06	−.13	−.17	−.04	−.14	−.13	−.02	−.13	−.45	−.05	−.15	−.32	−.04	−.16	−.29	−.02	−.15
	60	40	−.20	−.03	−.02	−.20	−.01	−.03	−.16	−.00	−.03	−.05	−.01	−.00	−.03	−.01	−.01	−.03	−.01	−.01	−.15	−.02	−.02	−.20	−.01	−.03	−.21	−.01	−.02
	100	40	−.28	−.02	−.01	−.14	−.01	−.02	−.15	−.01	−.02	−.02	−.00	−.00	−.04	−.01	−.00	−.01	−.00	−.00	−.23	−.01	−.02	−.14	−.01	−.02	−.20	−.00	−.02
	500	40	−.26	−.02	−.01	−.18	−.00	−.02	−.14	−.01	−.02	−.00	−.00	−.00	−.00	−.00	−.00	−.00	−.00	−.00	−.22	−.01	−.02	−.19	−.00	−.02	−.19	−.00	−.02

Note: ind_b = strength of between indirect effect; n_2 = number of groups; n_1 = group size; ρ = population interclass correlation coefficient; CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_{ab} = covariance of a and b pathways where 1 = −.113, 2 = 0, and 3 = .113.

negative. This was not surprising as the true value of the effect would be very small, −.023.

As in past studies, the sample size needed to reach acceptable power for the between indirect effects was lower for UMM than MSEM (Table 8). When the between indirect effect was small (ind_b = .01) and the cluster effect was small (ICC = .1), UMM reached acceptable power when the sample size was 500 groups. For the same conditions, power under MSEM was strikingly low, and in fact, it never reached above .2. When the between indirect effect was of medium strength (ind_b = .09) both UMM and MSEM required 500 groups to reach acceptable power. When the between indirect effect was large (ind_b = .36), UMM only required 60 groups to reach acceptable power, whereas MSEM required at least 100. This difference in UMM and MSEM also held when the cluster effect was large (ICC = .3).

Results for condition 2: Within-cluster mediation

CI coverage. When between pathways were set to 0, CI coverage for estimated within indirect effects were excellent (nearly all were within the acceptable range) for all cluster-level effects, sample sizes, and covariances of a and b (Table 9).

Bias. Relative percentage bias for the within indirect effects was again highest for the negative covariance conditions for all three methods, especially when cluster level effects of X, M, and Y were large (ICCs = .3; Table 10). However, it is important to note that UMM showed the lowest relative bias in within indirect effects, even in the negative covariance condition (only one cell had >10% bias outside of the smallest level-2 sample size). When a cluster-level effect only existed for the outcome (Y), the bias was related to the sample size and covariance, with higher sample sizes and stronger within pathways showing very low bias. This relationship held for all three methods. It also is important to note the direction of the relative bias for most conditions was negative, especially when the bias was large. This could indicate that the within indirect effects were being under-estimated when the between pathways were 0.

Power of within mediation. The power of within indirect effects was strongly related to the covariance of the a and b pathways (and thus, the strength of the effect) and sample size (Table 11). This relationship was the same for CMM, UMM, and MSEM. When the covariance was positive, and the subsequent within indirect effect the highest value in the simulation, adequate power was reached over all sample sizes and ICC's. When the covariance was set to 0, power to detect significant within indirect effects was dependent

Table 6. Relative bias for between indirect effects in 1-1-1 designs for UMM and MSEM strategies.

			UMM									MSEM								
			ind_b = .36			ind_b = .09			ind_b = .01			ind_b = .36			ind_b = .09			ind_b = .01		
cov_{ab}	n_2	n_1	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113
$\rho = .1$	30	20	−.42	−.44	−.45	−.13	−.10	−.08	1.31	1.71	1.93	−.32	−.25	−.18	−.26	−.10	−.05	−.27	−.29	−.14
	60	20	−.29	−.30	−.30	−.07	−.00	−.07	1.42	1.75	2.20	−.15	−.05	−.05	−.16	−.00	−.16	−.12	−.12	−.41
	100	20	−.30	−.29	−.29	−.05	−.02	−.06	1.40	1.79	2.20	−.16	−.05	−.08	−.15	−.01	−.15	−.02	−.22	−.31
	500	20	−.30	−.29	−.28	−.05	−.01	−.05	1.44	1.78	2.14	−.18	−.05	−.06	−.12	−.01	−.13	−.08	−.21	−.34
	30	40	−.37	−.35	−.36	−.09	−.07	−.09	−.68	−.90	1.25	−.38	−.24	−.16	−.20	−.05	−.05	−.22	−.29	−.83
	60	40	−.20	−.19	−.18	−.05	−.00	−.04	−.78	1.06	1.22	−.21	−.07	−.05	−.14	−.01	−.16	−.01	−.28	−.44
	100	40	−.18	−.18	−.18	−.02	−.01	−.03	−.81	1.02	1.22	−.20	−.07	−.06	−.12	−.01	−.14	−.09	−.33	−.47
	500	40	−.19	−.18	−.18	−.04	−.00	−.03	−.80	1.00	1.22	−.21	−.07	−.05	−.14	−.00	−.13	−.17	−.30	−.47
$\rho = .3$	30	20	−.28	−.30	−.30	−.06	−.03	−.04	−.35	−.53	−.42	−.33	−.24	−.19	−.13	−.06	−.11	−.26	−.21	−.31
	60	20	−.12	−.12	−.12	−.03	−.03	−.02	−.41	−.50	−.59	−.19	−.06	−.04	−.14	−.02	−.13	−.03	−.15	−.40
	100	20	−.11	−.11	−.11	−.03	−.01	−.03	−.44	−.54	−.63	−.17	−.05	−.06	−.14	−.02	−.13	−.04	−.21	−.40
	500	20	−.11	−.10	−.10	−.02	−.00	−.02	−.40	−.51	−.63	−.17	−.05	−.06	−.12	−.00	−.12	−.06	−.19	−.39
	30	40	−.24	−.26	−.25	−.09	−.04	−.07	−.24	−.21	−.22	−.35	−.24	−.14	−.21	−.06	−.03	−.06	−.19	−.54
	60	40	−.07	−.07	−.08	−.01	−.01	−.01	−.22	−.24	−.34	−.22	−.08	−.02	−.14	−.02	−.16	−.06	−.23	−.54
	100	40	−.06	−.06	−.06	−.01	−.02	−.02	−.19	−.29	−.33	−.21	−.07	−.06	−.14	−.02	−.15	−.06	−.25	−.55
	500	40	−.06	−.05	−.06	−.00	−.01	−.01	−.22	−.27	−.34	−.20	−.07	−.05	−.13	−.01	−.13	−.14	−.29	−.49

Note: ind_b = strength of between indirect effect; n_2 = number of groups; n_1 = group size; ρ = population interclass correlation coefficient; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_{ab} = covariance of a and b pathways.

Table 7. Power for within indirect effects in 1-1-1 designs for CMM, UMM, and MSEM (where a_w and b_w = .3).

			CMM						UMM						MSEM					
			ind_b = .36		ind_b = .09		ind_b = .01		ind_b = .36		ind_b = .09		ind_b = .01		ind_b = .36		ind_b = .09		ind_b = .01	
cov_{ab}	n_2	n_1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
$\rho = .1$	30	20	0.05	0.49	0.07	0.51	0.06	0.49	0.05	0.45	0.08	0.49	0.06	0.48	0.05	0.46	0.07	0.50	0.06	0.49
	60	20	0.07	0.81	0.06	0.79	0.07	0.78	0.09	0.76	0.08	0.76	0.09	0.76	0.07	0.79	0.06	0.79	0.06	0.78
	100	20	0.07	0.95	0.08	0.94	0.10	0.94	0.13	0.93	0.13	0.92	0.13	0.93	0.09	0.95	0.08	0.93	0.09	0.94
	500	20	0.24	11.0	0.31	11.0	0.37	11.0	0.51	11.0	0.52	11.0	0.52	11.0	0.29	11.0	0.32	11.0	0.30	11.0
	30	40	0.07	0.57	0.06	0.56	0.05	0.54	0.08	0.55	0.07	0.55	0.06	0.53	0.07	0.56	0.06	0.55	0.05	0.54
	60	40	0.08	0.84	0.08	0.84	0.08	0.84	0.10	0.82	0.09	0.83	0.10	0.83	0.08	0.84	0.08	0.84	0.08	0.85
	100	40	0.10	0.97	0.11	0.97	0.12	0.97	0.14	0.95	0.14	0.96	0.14	0.97	0.10	0.96	0.11	0.97	0.11	0.97
	500	40	0.39	11.0	0.47	11.0	0.50	11.0	0.59	11.0	0.59	11.0	0.60	11.0	0.43	11.0	0.47	11.0	0.46	11.0
$\rho = .3$	30	20	0.05	0.50	0.05	0.51	0.06	0.50	0.06	0.46	0.06	0.48	0.06	0.50	0.05	0.48	0.05	0.50	0.05	0.50
	60	20	0.05	0.82	0.08	0.81	0.08	0.79	0.09	0.77	0.10	0.76	0.10	0.77	0.06	0.81	0.07	0.80	0.07	0.80
	100	20	0.07	0.97	0.08	0.94	0.08	0.95	0.13	0.94	0.13	0.93	0.11	0.93	0.08	0.96	0.08	0.94	0.07	0.95
	500	20	0.17	11.0	0.28	11.0	0.35	11.0	0.54	11.0	0.51	11.0	0.53	11.0	0.24	11.0	0.28	11.0	0.26	11.0
	30	40	0.06	0.55	0.06	0.57	0.06	0.56	0.07	0.52	0.06	0.55	0.07	0.55	0.06	0.53	0.06	0.56	0.06	0.56
	60	40	0.08	0.87	0.07	0.85	0.08	0.86	0.10	0.84	0.09	0.83	0.10	0.85	0.08	0.87	0.07	0.85	0.08	0.86
	100	40	0.09	0.97	0.11	0.97	0.12	0.97	0.13	0.96	0.14	0.96	0.14	0.96	0.10	0.97	0.11	0.97	0.11	0.97
	500	40	0.37	11.0	0.45	11.0	0.49	11.0	0.58	11.0	0.58	11.0	0.59	11.0	0.40	11.0	0.45	11.0	0.43	11.0

Note: Power for covariance = .113 are not shown, as power was >.99 for levels of all other factors. ind_b = strength of between indirect effect; n_2 = number of groups; n_1 = group size; ρ = population interclass correlation coefficient; CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_{ab} = covariance of a and b pathways where 1 = −.113 or 2 = 0.

on sample size. For all three methods, adequate power was reached when at least 100 groups were used with 20 participants per group, or 60 groups with 40 participants per group. When covariance was negative, adequate power was never reached under the simulation conditions, for all three methods.

Type I Error of between mediation. In this simulation condition, only within mediation was simulated, and we explicitly set the between pathways to 0. Thus, we were able to examine the Type I Error rate for between indirect effects when UMM and MSEM were used to estimate the models. For MSEM, Type I Error rate was very low for all sample sizes and cluster effect levels (Table 12). However, for UMM, Type I Error rate increased with sample size. When the cluster level was small and only simulated for the outcome (Y) and 500 groups were used, between pathways were significant under UMM in over 90% of the replications. Although it still increased, this intensity of increase in Type I Error was not seen with 500 groups when the cluster-level effect was simulated for X, M, and Y together. When the cluster level effect was large for all three variables (ICCs = .3), the significance rates for the between indirect effects for UMM looked more similar to MSEM.

Results for condition 3: no mediation

The final simulation condition was used to evaluate Type I Error rates of both within and between effects

Table 8. Power for between indirect effects in 1-1-1 designs for UMM and MSEM strategies.

			UMM									MSEM								
			$ind_b = .36$			$ind_b = .09$			$ind_b = .01$			$ind_b = .36$			$ind_b = .09$			$ind_b = .01$		
cov_{ab}	n_2	n_1	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113	−.113	0	.113
$\rho = .1$	30	20	0.61	0.62	0.61	0.14	0.14	0.16	0.02	0.02	0.03	0.15	0.19	0.24	0.02	0.02	0.03	0.00	0.00	0.00
	60	20	0.95	0.96	0.95	0.37	0.41	0.41	0.05	0.04	0.05	0.57	0.65	0.71	0.05	0.07	0.10	0.00	0.01	0.00
	100	20	11.0	11.0	11.0	0.69	0.74	0.74	0.08	0.10	0.11	0.89	0.94	0.96	0.16	0.22	0.26	0.00	0.00	0.00
	500	20	11.0	11.0	11.0	11.0	11.0	11.0	0.91	0.95	0.97	11.0	11.0	11.0	0.99	11.0	11.0	0.07	0.08	0.08
	30	40	0.72	0.74	0.76	0.15	0.16	0.16	0.02	0.01	0.01	0.34	0.43	0.51	0.05	0.05	0.06	0.00	0.00	0.00
	60	40	0.98	0.98	0.99	0.39	0.41	0.44	0.02	0.02	0.03	0.76	0.84	0.92	0.13	0.15	0.19	0.00	0.00	0.00
	100	40	11.0	11.0	11.0	0.78	0.77	0.78	0.05	0.06	0.06	0.96	0.99	0.99	0.31	0.37	0.42	0.01	0.01	0.01
	500	40	11.0	11.0	11.0	11.0	11.0	11.0	0.79	0.84	0.88	11.0	11.0	11.0	11.0	11.0	11.0	0.16	0.18	0.20
$\rho = .3$	30	20	0.81	0.80	0.81	0.18	0.17	0.17	0.01	0.02	0.01	0.45	0.54	0.60	0.07	0.06	0.08	0.00	0.01	0.00
	60	20	0.99	0.99	0.99	0.42	0.46	0.46	0.02	0.01	0.02	0.86	0.93	0.96	0.15	0.20	0.23	0.00	0.00	0.00
	100	20	11.0	11.0	11.0	0.77	0.79	0.79	0.03	0.04	0.04	0.98	11.0	11.0	0.39	0.48	0.55	0.01	0.01	0.01
	500	20	11.0	11.0	11.0	11.0	11.0	11.0	0.62	0.65	0.72	11.0	11.0	11.0	11.0	11.0	11.0	0.19	0.22	0.25
	30	40	0.83	0.84	0.84	0.15	0.18	0.17	0.01	0.01	0.01	0.51	0.61	0.69	0.07	0.09	0.11	0.01	0.01	0.01
	60	40	0.99	0.99	0.99	0.44	0.47	0.45	0.01	0.01	0.01	0.86	0.94	0.97	0.20	0.26	0.30	0.01	0.01	0.01
	100	40	11.0	11.0	11.0	0.80	0.79	0.81	0.03	0.03	0.03	0.98	11.0	11.0	0.46	0.53	0.61	0.01	0.01	0.01
	500	40	11.0	11.0	11.0	11.0	11.0	11.0	0.51	0.55	0.59	11.0	11.0	11.0	11.0	11.0	11.0	0.26	0.29	0.33

Note: ind_b = strength of between indirect effect; n_2 = number of groups; n_1 = group size; ρ = population interclass correlation coefficient; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_{ab} = covariance of *a* and *b* pathways.

Table 9. Confidence interval coverage for within indirect effects in 1-1-1 designs for CMM, UMM, and MSEM when only within pathways are modeled (where a_w and $b_w = .3$).

			CMM			UMM			MSEM		
cov_{ab}	n_2	n_1	1	2	3	1	2	3	1	2	3
$\rho_{all} = .1$	30	20	0.93	0.91	0.81	0.93	0.91	0.84	0.92	0.91	0.82
	60	20	0.93	0.95	0.91	0.93	0.95	0.94	0.92	0.96	0.92
	100	20	0.93	0.94	0.92	0.94	0.95	0.93	0.92	0.94	0.93
	500	20	0.94	0.95	0.91	0.96	0.95	0.95	0.90	0.95	0.93
	30	40	0.92	0.93	0.85	0.93	0.93	0.87	0.92	0.93	0.86
	60	40	0.93	0.93	0.91	0.94	0.94	0.92	0.93	0.93	0.91
	100	40	0.95	0.94	0.92	0.95	0.94	0.93	0.94	0.94	0.92
	500	40	0.93	0.96	0.93	0.95	0.96	0.95	0.93	0.96	0.94
$\rho_{all} = .3$	30	20	0.91	0.93	0.81	0.93	0.93	0.84	0.91	0.93	0.82
	60	20	0.91	0.94	0.89	0.94	0.94	0.92	0.90	0.94	0.89
	100	20	0.92	0.94	0.90	0.95	0.94	0.92	0.91	0.94	0.91
	500	20	0.91	0.95	0.88	0.94	0.96	0.95	0.86	0.96	0.91
	30	40	0.92	0.93	0.85	0.93	0.93	0.87	0.91	0.93	0.86
	60	40	0.92	0.95	0.90	0.92	0.95	0.91	0.92	0.95	0.90
	100	40	0.94	0.95	0.92	0.94	0.95	0.93	0.94	0.95	0.93
	500	40	0.93	0.94	0.94	0.93	0.95	0.96	0.91	0.95	0.94
$\rho_y = .1$	30	20	0.93	0.93	0.79	0.93	0.93	0.79	0.93	0.93	0.80
	60	20	0.94	0.95	0.93	0.94	0.94	0.93	0.95	0.95	0.93
	100	20	0.95	0.95	0.94	0.95	0.94	0.94	0.95	0.95	0.94
	500	20	0.96	0.95	0.94	0.96	0.95	0.94	0.96	0.95	0.94
	30	40	0.92	0.92	0.84	0.93	0.93	0.82	0.92	0.93	0.83
	60	40	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
	100	40	0.94	0.94	0.94	0.93	0.94	0.94	0.94	0.94	0.94
	500	40	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
$\rho_y = .3$	30	20	0.96	0.92	0.85	0.96	0.92	0.85	0.97	0.92	0.85
	60	20	0.94	0.93	0.90	0.93	0.93	0.90	0.94	0.93	0.90
	100	20	0.96	0.93	0.91	0.95	0.94	0.91	0.95	0.93	0.91
	500	20	0.93	0.95	0.94	0.93	0.95	0.93	0.93	0.95	0.94
	30	40	0.94	0.81	0.78	0.94	0.82	0.81	0.93	0.82	0.78
	60	40	0.92	0.94	0.85	0.93	0.94	0.85	0.92	0.94	0.85
	100	40	0.93	0.96	0.94	0.93	0.96	0.93	0.93	0.96	0.94
	500	40	0.96	0.92	0.96	0.95	0.92	0.96	0.95	0.91	0.95

Note: n_2 = number of groups; n_1 = group size; ρ_{all} = population interclass correlation coefficient for outcome (*Y*), predictor (*X*), and mediator (*M*); ρ_y = population interclass correlation coefficient for outcome (*Y*) where ICC's for *X* and *M* are 0; CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_{ab} = covariance of *a* and *b* pathways where 1 = −.113, 2 = 0, and 3 = .113.

for CMM, UMM, and MSEM (Table 13). For these replications, all of the mediation pathways were set to 0. For between indirect effects, Type I Error rates were low (<.02) for both UMM and MSEM over all simulated cluster levels and sample sizes. For within indirect effects, all three methods showed an inverse relationship between error rates and sample size. As the sample size increased, Type I Error rates for

Table 10. Relative bias for within indirect effects in 1-1-1 designs for CMM, UMM, and MSEM strategies when only within pathways were modeled (where a_w and $b_w = .3$).

			CMM			UMM			MSEM		
cov_{ab}	n_2	n_1	1	2	3	1	2	3	1	2	3
$\rho_{all} = .1$	30	20	−0.48	−0.09	−0.19	−0.33	−0.07	−0.16	−0.60	−0.07	−0.18
	60	20	−0.24	−0.02	−0.05	−0.06	0.00	−0.02	−0.35	0.01	−0.04
	100	20	−0.25	−0.01	−0.03	−0.07	0.01	0.00	−0.36	0.02	−0.02
	500	20	−0.20	−0.01	−0.04	0.00	0.00	0.00	−0.31	0.01	−0.03
	30	40	−0.25	−0.06	−0.14	−0.16	−0.05	−0.12	−0.31	−0.05	−0.14
	60	40	−0.13	0.00	−0.03	−0.02	0.01	−0.01	−0.18	0.01	−0.02
	100	40	−0.08	0.00	−0.02	0.03	0.01	0.00	−0.14	0.02	−0.02
	500	40	−0.10	−0.01	−0.02	0.02	0.00	0.00	−0.16	0.00	−0.01
$\rho_{all} = .3$	30	20	−0.50	−0.11	−0.20	−0.30	−0.09	−0.16	−0.65	−0.08	−0.19
	60	20	−0.31	−0.01	−0.07	−0.08	0.01	−0.02	−0.45	0.02	−0.06
	100	20	−0.25	−0.03	−0.05	−0.01	−0.01	−0.01	−0.39	0.00	−0.04
	500	20	−0.24	−0.01	−0.05	0.01	0.01	0.00	−0.38	0.01	−0.04
	30	40	−0.36	−0.06	−0.15	−0.25	−0.05	−0.13	−0.43	−0.04	−0.15
	60	40	−0.12	−0.02	−0.03	0.00	−0.01	−0.01	−0.19	−0.01	−0.03
	100	40	−0.11	0.01	−0.03	0.02	0.02	0.00	−0.18	0.03	−0.02
	500	40	−0.14	−0.01	−0.02	−0.01	0.00	0.00	−0.21	0.01	−0.02
$\rho_y = .1$	30	20	−0.25	−0.11	−0.19	−0.27	−0.10	−0.19	−0.26	−0.11	−0.19
	60	20	−0.02	0.02	−0.02	−0.03	0.02	−0.02	−0.03	0.02	−0.02
	100	20	−0.03	0.00	0.00	−0.04	0.00	0.00	−0.03	0.00	0.00
	500	20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	30	40	−0.16	−0.12	−0.18	−0.17	−0.12	−0.17	−0.17	−0.12	−0.18
	60	40	−0.07	0.00	−0.02	−0.09	0.00	−0.02	−0.08	0.00	−0.02
	100	40	0.02	0.00	−0.02	0.02	0.00	−0.02	0.02	0.00	−0.02
	500	40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\rho_y = .3$	30	20	−0.44	−0.21	−0.20	−0.49	−0.20	−0.20	−0.47	−0.20	−0.20
	60	20	−0.02	−0.06	−0.06	−0.05	−0.05	−0.05	−0.03	−0.06	−0.06
	100	20	0.06	0.02	−0.03	0.05	0.02	−0.02	0.06	0.02	−0.02
	500	20	0.03	−0.01	0.00	0.02	0.00	0.00	0.02	−0.01	0.00
	30	40	−0.08	−0.21	−0.20	−0.11	−0.20	−0.20	−0.10	−0.21	−0.20
	60	40	0.18	−0.08	−0.08	0.16	−0.07	−0.08	0.17	−0.08	−0.08
	100	40	−0.01	0.00	−0.01	−0.02	0.00	−0.01	−0.01	0.00	−0.01
	500	40	−0.01	−0.01	0.00	−0.02	−0.01	0.00	−0.01	−0.01	0.00

Note: n_2 = number of groups; n_1 = group size; ρ_{all} = population interclass correlation coefficient for outcome (*Y*), predictor (*X*), and mediator (*M*); ρ_y = population interclass correlation coefficient for outcome (*Y*) where ICC's for *X* and *M* are 0; CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_{ab} = covariance of *a* and *b* pathways where 1 = −.113, 2 = 0, and 3 = .113.

within indirect effects generally decreased. This relationship held over all the cluster-level effects.

Relative bias in standard error estimation

For the within indirect effects, bias in the standard errors was negligible for all three methods; the absolute value of the mean relative bias was <.01 (standard deviation's all equaled .15) for CMM, UMM, and MSEM (results not shown). However, relative bias in the standard errors of between indirect effects showed more variability and results were different for the three conditions explored in the simulation. In the full mediation condition (i.e., ICCs for *X*, *M*, and *Y* were simulated and the between indirect effects were nonzero), relative bias in the standard errors for the between indirect effects was generally negligible (<.05) for both UMM and MSEM (Table 14). There was one notable exception: when the between indirect effect was very small (.01), MSEM showed a mean bias of .17 when sample size was lowest (compared to −.06 for UMM under the same conditions).

For the other two conditions, relative bias in the standard errors was generally larger than in the full mediation condition, and larger still when estimated with MSEM versus UMM. For the within-only mediation condition (where ICC's were varied but between indirect effects were zero), relative bias in the standard errors was dependent on both the estimation method and the ICCs (Table 14). For example, even when sample sizes were large, MSEM had higher relative bias (.11 vs <.01 for UMM) when $\rho_{all} = .1$. When ICCs for all variables were larger ($\rho_{all} = .3$), the methods looked generally similar but the relative bias in the standard errors were still above .10 for most sample sizes. When ICCs were only simulated for *Y*, standard errors were notably overestimated when using MSEM, even at the higher levels of sample size included in the simulation. This was less of an issue for UMM; when $\rho_y = .1$, bias was negligible (<.02) for the standard errors when 500 groups were simulated. For the final simulation condition (no mediation), the standard errors were overestimated (relative bias was large and positive) for all conditions (Table 14).

Table 11. Power for within indirect effects in 1-1-1 designs for CMM, UMM, and MSEM strategies when only within pathways are modeled (where a_w and $b_w = .3$).

			CMM			UMM			MSEM		
cov_{ab}	n_2	n_1	1	2	3	1	2	3	1	2	3
$\rho_{all} = .1$	30	20	0.06	0.51	0.99	0.06	0.50	0.99	0.05	0.51	0.99
	60	20	0.08	0.79	1.00	0.09	0.78	1.00	0.08	0.80	1.00
	100	20	0.09	0.95	1.00	0.12	0.94	1.00	0.07	0.95	1.00
	500	20	0.38	1.00	1.00	0.52	1.00	1.00	0.28	1.00	1.00
	30	40	0.07	0.54	1.00	0.07	0.54	1.00	0.06	0.55	1.00
	60	40	0.09	0.84	1.00	0.11	0.83	1.00	0.07	0.85	1.00
	100	40	0.14	0.96	1.00	0.16	0.96	1.00	0.12	0.96	1.00
	500	40	0.52	1.00	1.00	0.60	1.00	1.00	0.46	1.00	1.00
$\rho_{all} = .3$	30	20	0.05	0.49	0.99	0.05	0.48	0.99	0.05	0.49	0.99
	60	20	0.07	0.80	1.00	0.09	0.77	1.00	0.06	0.81	1.00
	100	20	0.11	0.95	1.00	0.14	0.93	1.00	0.09	0.95	1.00
	500	20	0.36	1.00	1.00	0.53	1.00	1.00	0.23	1.00	1.00
	30	40	0.05	0.54	1.00	0.06	0.53	1.00	0.05	0.55	1.00
	60	40	0.08	0.84	1.00	0.09	0.83	1.00	0.07	0.84	1.00
	100	40	0.12	0.97	1.00	0.14	0.97	1.00	0.11	0.97	1.00
	500	40	0.49	1.00	1.00	0.58	1.00	1.00	0.41	1.00	1.00
$\rho_y = .1$	30	20	0.07	0.44	0.99	0.07	0.46	0.98	0.07	0.44	0.98
	60	20	0.09	0.79	1.00	0.08	0.78	1.00	0.08	0.79	1.00
	100	20	0.13	0.94	1.00	0.13	0.94	1.00	0.13	0.94	1.00
	500	20	0.53	1.00	1.00	0.51	1.00	1.00	0.53	1.00	1.00
	30	40	0.09	0.51	1.00	0.09	0.50	1.00	0.08	0.50	1.00
	60	40	0.08	0.84	1.00	0.09	0.84	1.00	0.08	0.84	1.00
	100	40	0.15	0.97	1.00	0.15	0.97	1.00	0.15	0.97	1.00
	500	40	0.59	1.00	1.00	0.59	1.00	1.00	0.59	1.00	1.00
$\rho_y = .3$	30	20	0.03	0.39	0.99	0.03	0.40	1.00	0.03	0.38	0.99
	60	20	0.08	0.72	1.00	0.09	0.73	1.00	0.08	0.72	1.00
	100	20	0.12	0.94	1.00	0.11	0.93	1.00	0.11	0.94	1.00
	500	20	0.54	1.00	1.00	0.53	1.00	1.00	0.53	1.00	1.00
	30	40	0.03	0.43	1.00	0.03	0.43	1.00	0.03	0.43	1.00
	60	40	0.15	0.75	1.00	0.15	0.75	1.00	0.15	0.76	1.00
	100	40	0.15	0.98	1.00	0.15	0.98	1.00	0.15	0.98	1.00
	500	40	0.56	1.00	1.00	0.55	1.00	1.00	0.56	1.00	1.00

Note: n_2 = number of groups; n_1 = group size; ρ_{all} = population interclass correlation coefficient for outcome (*Y*), predictor (*X*), and mediator (*M*); ρ_y = population interclass correlation coefficient for outcome (*Y*) where ICC's for *X* and *M* are 0; CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; cov_{ab} = covariance of *a* and *b* pathways where 1 = –.113, 2 = 0, and 3 = .113.

Empirical example

Student engagement, classroom climate, and student achievement: a 1-1-1 mediation model

For our empirical example, previously collected data from a large study that assessed middle and high school student engagement in math and science was analyzed in the context of mediation (Wang, Fredricks, Ye, Hofkens, & Linn, 2016). Developing a student's interest in math and science has far reaching consequences like the selection of a college major or obtaining a successful career in science, technology, engineering, and mathematics (STEM; Wang & Degol, 2014). Aspects of classroom climate, like instructional practices, interactions with teachers, and classroom organization, create opportunities for adolescents to engage in a variety of math-related activities and build relationships in the classroom (Eccles, Wigfield, & Schiefele, 1998). Student engagement is a complex topic, and researchers have become more and more interested in how engagement interacts with other variables to influence academic success. Researchers have found that student engagement influences individual math learning, and it is conceptualized as malleable and responsive to variations in classroom and teacher characteristics. In this empirical example, we explored the differences and similarities between CMM, UMM, and MSEM when estimating if student engagement (*M*) mediated the relationship between classroom climate (*X*) and math achievement (*Y*).

Methods

Sample 1. The sample was recruited from four high schools (grades 8–12) located in a socioeconomically and ethnically diverse community in western Pennsylvania. To avoid the impact of missing data on model performance, students with missing values on any of the three variables were deleted from analysis. The total sample included 783 students nested within 173 math classrooms (n_2). Within each classroom, the number of students

Table 12. Type I Error for between indirect effects in 1-1-1 designs for CMM, UMM, and MSEM when only within pathways were simulated and between pathways were set to 0 (condition 2).

			UMM			MSEM		
	n_2	n_1	$cov_{ab} = -.113$	$cov_{ab} = 0$	$cov_{ab} = .113$	$cov_{ab} = -.113$	$cov_{ab} = 0$	$cov_{ab} = .113$
$\rho_{all} = .1$	30	20	.00	.01	.00	.00	.00	.00
	60	20	.01	.01	.02	.00	.00	.00
	100	20	.01	.02	.01	.00	.00	.00
	500	20	.30	.41	.46	.00	.00	.00
	30	40	.01	.00	.01	.00	.00	.00
	60	40	.00	.01	.00	.00	.00	.00
	100	40	.00	.00	.01	.00	.00	.00
	500	40	.03	.05	.08	.00	.00	.00
$\rho_{all} = .3$	30	20	.01	.00	.00	.00	.00	.00
	60	20	.00	.00	.00	.00	.00	.00
	100	20	.00	.00	.00	.00	.00	.00
	500	20	.00	.01	.01	.00	.00	.00
	30	40	.00	.00	.00	.00	.00	.00
	60	40	.00	.00	.00	.00	.00	.00
	100	40	.00	.00	.00	.00	.00	.00
	500	40	.00	.00	.00	.00	.00	.00
$\rho_y = .1$	30	20	.03	.02	.01	.01	.02	.01
	60	20	.09	.10	.09	.01	.01	.01
	100	20	.20	.23	.23	.00	.00	.01
	500	20	.96	.99	.99	.00	.00	.00
	30	40	.02	.01	.00	.01	.01	.01
	60	40	.04	.04	.03	.01	.01	.01
	100	40	.12	.15	.14	.01	.00	.01
	500	40	.84	.91	.96	.00	.00	.00
$\rho_y = .3$	30	20	.00	.02	.00	.01	.01	.00
	60	20	.02	.02	.01	.00	.01	.01
	100	20	.07	.04	.05	.01	.01	.01
	500	20	.65	.74	.83	.00	.00	.00
	30	40	.00	.00	.00	.00	.00	.00
	60	40	.01	.00	.00	.00	.00	.01
	100	40	.02	.02	.01	.00	.00	.01
	500	40	.37	.46	.55	.00	.01	.01

Note: n_2 = number of groups; n_1 = group size; ρ_{all} = population interclass correlation coefficient for outcome (*Y*), predictor (*X*), and mediator (*M*); ρ_y = population interclass correlation coefficient for outcome (*Y*) where ICC's for *X* and *M* are 0; CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling.

Table 13. Type I error rates for indirect effects in 1-1-1 designs for CMM, UMM, and MSEM when all pathways were set to 0 ("no mediation condition").

			CMM				UMM				MSEM			
	n_2	n_1	$\rho_{all} = .1$	$\rho_{all} = .3$	$\rho_y = .1$	$\rho_y = .3$	$\rho_y = .1$	$\rho_y = .3$	$\rho_y = .1$	$\rho_y = .3$	$\rho_y = .1$	$\rho_y = .3$	$\rho_y = .1$	$\rho_y = .3$
Within indirect effects	30	20	.07	.07	.05	.05	.07	.07	.05	.05	.06	.06	.05	.05
	60	20	.05	.06	.06	.06	.06	.06	.06	.06	.05	.05	.06	.06
	100	20	.05	.04	.05	.07	.05	.05	.05	.07	.05	.04	.05	.07
	500	20	.05	.06	.06	.04	.04	.06	.06	.04	.04	.06	.06	.04
	30	40	.06	.06	.07	.05	.06	.06	.07	.06	.05	.06	.06	.05
	60	40	.05	.05	.06	.06	.05	.05	.06	.06	.05	.05	.06	.06
	100	40	.06	.06	.07	.06	.06	.05	.07	.06	.06	.05	.07	.06
	500	40	.05	.06	.05	.06	.05	.06	.04	.06	.05	.06	.05	.06
Between indirect effects	30	20					.00	.00	.00	.00	.00	.00	.02	.01
	60	20					.00	.00	.00	.00	.00	.00	.00	.00
	100	20					.00	.00	.00	.00	.00	.00	.00	.00
	500	20					.00	.00	.00	.00	.00	.00	.00	.00
	30	40					.00	.00	.00	.01	.00	.00	.01	.01
	60	40					.00	.00	.00	.00	.00	.00	.00	.00
	100	40					.00	.00	.00	.00	.00	.00	.00	.00
	500	40					.00	.00	.00	.00	.00	.00	.00	.00

Note: n_2 = number of groups; n_1 = group size; ρ_{all} = population interclass correlation coefficient for outcome (*Y*), mediator (*M*), and predictor (*X*); ρ_y = population interclass correlation coefficient for outcome (*Y*), ICCs for *M* and *X* set to 0; CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling.

ranged from 1 to 20, with an overall small average cluster size (*mean* $n_1 = 4.5$). The sample was evenly split on gender (female =49%), and the ethnic breakdown was 24% Black/African American and 56% White/Caucasian American. Sixty percent of the students were receiving free or reduced price lunch.

It is also important to note that for the purposes of this example, other student and classroom level

Table 14. Relative bias in standard error estimates for the between indirect effect when using UMM or MSEM over all simulation replications.

				UMM					MSEM				
	n_2	n_1	*ind*b=	.36	.09	.01	.0	.0*	.36	.09	.01	.0	.0*
$\rho_{all} = .1$	30	20		−0.04	−0.04	−0.06	0.01	0.15	0.06	0.01	0.17	0.15	0.13
	60	20		−0.02	−0.05	−0.04	−0.02	0.19	−0.02	−0.04	0.06	0.09	0.11
	100	20		−0.04	−0.04	−0.03	0.00	0.20	−0.04	−0.01	0.07	0.18	0.14
	500	20		−0.01	−0.01	−0.01	−0.02	0.26	0.00	−0.01	0.00	0.16	0.28
	30	40		−0.06	−0.05	−0.02	0.06	0.13	−0.05	−0.05	0.03	0.08	0.08
	60	40		−0.04	−0.04	−0.03	0.07	0.17	−0.03	−0.05	0.04	0.15	0.15
	100	40		−0.03	−0.03	−0.03	0.07	0.20	−0.02	−0.03	0.02	0.19	0.24
	500	40		−0.01	0.00	0.00	0.00	0.27	−0.02	0.01	0.00	0.11	0.27
$\rho_{all} = .3$	30	20		−0.06	−0.05	0.00	0.10	0.14	−0.05	−0.04	0.04	0.12	0.05
	60	20		−0.02	−0.04	−0.02	0.13	0.18	−0.01	−0.02	0.04	0.18	0.18
	100	20		−0.02	−0.04	−0.03	0.11	0.23	−0.02	−0.05	0.02	0.19	0.23
	500	20		0.01	−0.01	0.00	0.04	0.23	0.00	−0.01	0.00	0.14	0.24
	30	40		−0.03	−0.06	−0.01	0.09	0.12	−0.06	−0.06	0.00	0.10	0.13
	60	40		−0.03	−0.04	−0.02	0.16	0.19	−0.04	−0.05	−0.01	0.14	0.21
	100	40		−0.04	−0.02	−0.02	0.16	0.20	−0.03	−0.02	0.00	0.14	0.21
	500	40		−0.01	−0.01	0.01	0.08	0.26	−0.01	−0.02	0.00	0.07	0.20
$\rho_y = .1$	30	20					0.14	0.29				0.86	0.82
	60	20					0.02	0.18				0.72	1.25
	100	20					−0.01	0.17				0.71	1.21
	500	20					0.00	0.27				0.54	1.13
	30	40					0.24	0.46				1.07	1.01
	60	40					0.14	0.10				0.88	1.10
	100	40					0.05	0.21				0.69	1.00
	500	40					0.02	0.23				0.63	0.88
$\rho_y = .3$	30	20					0.48	0.52				1.07	1.18
	60	20					0.20	0.32				0.87	1.11
	100	20					0.13	0.09				0.74	0.80
	500	20					0.10	0.36				0.82	0.46
	30	40					0.80	1.17				1.47	0.96
	60	40					0.59	0.61				1.35	1.31
	100	40					0.30	0.25				1.35	1.06
	500	40					0.26	0.32				0.76	0.59

Note: ind_b = between indirect effect; ρ_{all} = population interclass correlation coefficient for outcome (*Y*), mediator (*M*), and predictor (*X*); ρ_y = population interclass correlation coefficient for outcome (*Y*), ICC's for *M* and *X* set to 0; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling. *Within pathways also set to be 0 (*a* and *b* and the covariance were set to be 0).

covariates (including the possible difference among the high schools) were ignored. This limited approach is not recommended for applied studies, where researchers should fully consider potential covariates for inclusion in the final model. Although this increases model complexity, controlling for supplementary variables can better reflect reality and explain additional variability.

Sample 2. Since the level-1 sample size (i.e., cluster size) was very small in the initial sample, we also selected a subset of larger classrooms that included data on 10 or more students. Although the overall resulting sample size was smaller (number of total students =276; $n_2 = 21$), we wanted to compare results using a larger level-1 sample size (*mean* $n_1 = 13.1$).

Measures. The student's view of classroom climate (*X*) was measured as a composite variable using seven items rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Items included; "My math teacher praises students' effort and hard work rather than high grades," "My math teacher challenges students to do their very best," and "My math teacher believes that mistakes are okay as long as students are learning from their mistakes." As cluster size was small, internal consistency of the composite score was calculated using multilevel *Cronbach's* α following Geldhof, Preacher, & Zyphur (2014). Both within and between *Cronbach's* α were excellent (.92 and .98, respectively).

Student engagement in math learning (*M*) was measured using a multidimensional scale that assessed students on behavioral (six items), cognitive (four items), and emotional engagement (five items). Items were answered on a 5-point Likert scale, and were selected from a larger item pool (as described in Wang et al., 2016). Items included, "I keep trying even if something is hard" (behavioral engagement), "I try to understand mistakes when I get something wrong" (cognitive engagement), and "I look forward to math class" (emotional engagement). For each student, overall engagement scores were calculated as the mean of the final 15 items. Internal consistency of the scores was excellent (within and between *Cronbach's* $\alpha = .94$ and .96).

Student achievement (*Y*) was reflected by the final grade received in the student's current math course. Grades were reported on a 100-point scale. The mean math grade for the full sample was 80.10 (SD = 12.90). Before fitting the multilevel mediation

Table 15. Results of Empirical Example: 1-1-1 Mediation with CMM, UMM, and MSEM strategies.

	Sample 1			Sample 2		
	CMM *Estimate (SE)*	UMM *Estimate (SE)*	MSEM *Estimate (SE)*	CMM *Estimate (SE)*	UMM *Estimate (SE)*	MSEM *Estimate (SE)*
Indirect Effects						
Within	.16 (.03)***	.14 (.04)***	.14 (.03)***	.25 (.09)**	.23 (.08)**	.23 (.08)**
Between		.26 (.06)***	4.87 (10.94)		.65 (.15)***	1.73 (.51)***
Path *a* *Climate→Engagement*						
Within	.57 (.04)***	.57 (.05)***	.55 (.04)***	.57 (.07)***	.57 (.08)***	.57 (.07)***
Variance	.06 (.02)**	.09 (.03)**	.06 (.02)**	.04 (.02)	.05 (.03)	.04 (.02)
Between		.55 (.07)***	.22 (.14)		.58 (.12)***	.67 (.18)***
Path *b* Engagement→Course Grade						
Within	.27 (.05)***	.24 (.05)***	.24 (.05)***	.40 (.12)**	.38 (.12)**	.38 (.13)**
Variance	.02 (02)	.02 (.03)	.02 (.02)	.06 (.07)	.07 (.08)	.05 (.07)
Between		.47 (.10)***	6.09 (14.17)		1.11 (.25)***	2.59 (1.02)*
Path *c′* Climate→Course Grade						
Within	.04 (.04)	.02 (.05)	.02 (.05)	−.14 (.11)	−.13 (.11)	−.13 (.11)
Variance	.01(.02)	.01 (.02)	.01 (.02)	.03 (.04)	.04 (.05)	.03 (.04)
Between		.08 (.08)	−.71 (3.33)		−.38 (.22)	−.11 (.56)

Note: Sample 1 ($N = 783$; $n_2 = 174$; *average* $n_1 = 4.5$) and Sample 2 ($N = 276$; $n_2 = 21$; *average* $n_1 = 13.4$), CMM = conflated multilevel modeling; UMM = unconflated multilevel modeling; MSEM = multilevel structural equation modeling; *$p < .05$, **$p < .01$, ***$p < .001$.

models, all variables were standardized by subtracting the grand mean and dividing by square root of the total variance (although it is important to note such standardization does not produce standardized path coefficients). For UMM, student engagement and perception of classroom management were classroom-mean centered.

Justification for design. All three of the variables were measured at the student level, making a 1-1-1 design appropriate. We also hypothesized that there would be both within-classroom and between-classroom mediation. When student perception of classroom climate and student engagement in math learning are aggregated at the classroom level, they theoretically represent characteristics of one math teacher who, in turn, can directly impact classroom math achievement.

Statistical analysis. *Mplus* 7.3 (Muthén & Muthén, 1998–2017) was used to estimate the 1-1-1 design using CMM, UMM, and MSEM. Syntax was the same as that used in the simulation study and was adopted from the supplemental materials provided for Preacher et al.'s (2010) and (2011) papers. Analysis was repeated on both the full sample and the smaller subset of larger classrooms.

Results of the empirical example

All three methods converged with no estimation error. Cluster-level effects were small for student ratings of classroom climate ($ICC_X = .15$) and student engagement ($ICC_M = .09$), but large for student achievement ($ICC_Y = .35$). For the full sample (Sample 1), there was almost no difference in terms of estimated within indirect effects between methods, with CMM estimates being slightly higher (all results reported in Table 15). This was not entirely unexpected, as CMM conflates the within- and between pathways. However, for between classroom pathways, there were striking differences in estimation when using UMM and MSEM. Most saliently, UMM identified the between indirect effects as statistically significant, while MSEM did not. When using MSEM, both the between pathways of *a* and *b* were nonsignificant, and their standard errors much larger than those estimated by UMM. It is important to again note, that even though the sample size was relatively large ($N = 783$) in this example, the average cluster size was small ($n_1 = 4.5$). For MSEM, the estimated between-classroom correlation was .75 for achievement (Y) and engagement (M), .56 for achievement (Y) and classroom climate (X), and .91 for engagement (M) and climate (X).

In the smaller sample, again, there were little differences among three methods regarding the within indirect effects (Table 15). The results for the between indirect effects were more similar for this group of students, as both MSEM and UMM found them to be statistically significant. However, the estimate and standard errors were still larger for MSEM. We were again limited by our sample, as an average cluster size of 13.14 is likely still too small to provide reliable estimates of population cluster means. However, an initial pattern was suggested that when cluster size gets larger, the estimates from MSEM and UMM become more equivalent. For this sample, the estimated between-classroom correlation was .78 for *Y* and *M*, .26 for *Y* and *X*, and .78 for *M* and *X*.

Discussion

Researchers who are looking to estimate mediation in multilevel data have three potential methods to choose from, CMM, UMM, and MSEM. However, there are limited resources to help choose the appropriate method for specific data. Currently, we know that for 2-1-1 designs with fixed within-cluster slopes, MSEM requires larger sample sizes than UMM (McNeish, 2017), but when sample size requirements are met, MSEM outperformed UMM and CMM over a number of data conditions and on multiple outcomes (Preacher et al., 2011). In our simulation study, we compared the performance of the three methods when estimating a 1-1-1 design under varying conditions that included sample sizes, cluster effects, and pathway strengths.

Our study was novel in a number of ways. To our knowledge, this is the first time the performance of MSEM has been compared directly to UMM and CMM in the estimation of a 1-1-1 design. Past researchers have focused on comparing these methods in 2-1-1 designs (McNeish, 2017; Preacher et al., 2010, 2011). The 2-1-1 design has interesting properties, as it can still be fit using CMM, UMM, and MSEM, but the indirect effect only exists at the between level since the predictor is only measured at level-2. In contrast, by exploring the 1-1-1 design, we were able to evaluate the estimates of both between and within indirect effects, and by varying the strength of each pathway, we also could explore performance when the indirect effects were unbalanced.

It also is important to note that the addition of random slopes in our simulation increases the complexity of the model, as it allows the direct and indirect effects to vary across clusters. Current comparisons of multilevel mediation models focus primarily on fixed within-cluster slopes (McNeish, 2017; Preacher et al., 2010, 2011). Allowing for variation or heterogeneity in the within-cluster slopes of a and b paths at level-2 can be of great interest to applied researchers and can imply the existence of moderating variables (Bauer et al., 2006). In the supplemental materials from their 2011 study, Preacher et al. included results from a 2-1-1 design with random slopes of the 1-1 linkage (b pathway). However, a 1-1-1 design allows exploration of randomness of slopes for both the a and b pathways. Finally, in our simulation study, we also explored three mediation conditions. The first was analogous to full multilevel mediation, with nonzero between and within pathways. The second condition simulated nonzero within pathways, but the between pathways were set to 0. The third condition was a "no mediation" condition with both between and within pathways set to 0. These three conditions provided us with insight into the performance of the methods over a wide range of data conditions, and allowed us to verify if the existence of a between indirect effect was related to performance of the methods. Our main goal was to examine conditions where CMM or UMM might perform as well as, if not inferior to MSEM.

Choosing a method

In our simulation study results, there were two instances when UMM was preferred over MSEM in a 1-1-1 design. First, UMM required lower sample sizes in overall conditions to reach the same power for detecting nonzero between indirect effects when compared to MSEM. For example, when between indirect pathways were larger than the within pathways ($\text{ind}_b = .36$; $\text{ind}_w = .09$) and ICCs small (.1), UMM showed power of around .96 with only 60 groups and 20 participants per group, while MSEM showed a power of .65 under the same conditions. These results are similar to two past studies when evaluating the methods under a 2-1-1 design. In Preacher and colleagues 2011 paper, they found that when ICCs were small (.1), power for between indirect effects for 50 groups with 20 participants per group was .675 when using UMM and only .414 when using MSEM (under the same conditions). In 2017, McNeish also explored the performance of UMM and MSEM, but specifically simulated low and unbalanced sample sizes. In a condition that is roughly equivalent to the abovementioned example from our study (medium effect sizes on Y, ICCs of .2), McNeish (2017) found that 50 clusters with unbalanced clusters of 7–14 participants per cluster were required for MLM to reach detection rates of around .99 for nonzero between indirect effects, while detection rates for MSEM were only .65.

The second condition when UMM was preferred over MSEM was when the covariance of a and b pathways was negative. In these conditions, relative bias for the within indirect effect was not as high for UMM when compared to CMM or MSEM. Including negative covariance to affect the strength of the within indirect effect was novel; as of publication, no other studies used covariance as a simulation condition when comparing the three methods. Negative covariance of the a and b pathways could occur when the two parameters are inversely related; as the pathway from X to M (a) increases, the pathway from M to Y (b) would decrease. The effect of negative covariance on relative bias is likely due to its influence on the total strength of the indirect effect. When we simulated data with negative covariance (–.113), it decreased the size of the population within indirect effect from .09 to –.023, a probable reason leading to large relative bias. Thus, based on our results, it follows that if either a very small within indirect effect is expected or negative covariance of the pathways could suppress the within indirect effect, UMM will likely result in lower bias than MSEM or CMM. However, it is also important to note that UMM did not show an improvement over the other two methods in the same conditions regarding power or CI coverage.

Despite these two conditions when UMM was preferred over MSEM, there were a number of conditions where MSEM was preferred. For between indirect effects, CI coverage rates were higher than UMM when the between pathways were larger than the within pathways, and relative bias was lower than UMM when the ICCs and the strength of indirect effect were smaller. These results replicate similar findings in previously published papers comparing these methods when estimating 2-1-1 designs. In Preacher and colleagues 2011 paper, they found that MSEM had much better CI coverage when compared to UMM when the between *b* pathway was .5 and the within *b* pathway was .2. As in our study, this advantage of MSEM held over all levels of ICCs (range = .05–.4). They also saw better relative bias for MSEM compared to UMM over all their simulation conditions, although in their study, the advantage of MSEM was much more drastic than in our study. This is likely due to the calculation of bias for between indirect effects in a 2-1-1 design (Preacher et al., 2011).

When applied researchers are only modeling the multilevel data structure to control for confounding effects, but are not explicitly interested in estimating separate between and within pathways, CMM is likely an acceptable alternative to UMM or MSEM. In our simulation study, CI coverage for CMM was comparable to both UMM and MSEM for within indirect effects for both mediation conditions (true multilevel mediation or within-mediation only). Relative bias of the within indirect effects was also comparable to UMM and MSEM, and very low ($\leq$5%), for conditions when the covariance was non-negative and group sizes were $\geq$60. Power for nonzero within indirect effects reached acceptable levels at relatively low sample sizes (around 60 groups for true multilevel mediation, and around 100 groups for within-mediation only), but again, this only held for non-negative covariance conditions. Finally, Type I Error rates when the within indirect effects were zero were also similar to UMM and MSEM.

Is the existence of between-cluster effects related to performance of CMM, UMM, or MSEM?

Surprisingly, we found only one major difference in method performance when multilevel mediation was simulated only for the within pathways. Specifically, we found that Type I Error rates for the *between* pathways were lower when using MSEM compared to UMM. More problematically, but not entirely surprisingly, the Type I Error rates of the between indirect effects for UMM increased with sample size. This was likely related to the standard errors of the between indirect effect, which, in our simulation study, tended to be overestimated when using MSEM (versus UMM), specifically when the between pathways were 0, ICCs were large, and sample sizes were high. Our interpretation is that UMM is more sensitive to ICC when estimating the between indirect effects, and that if cluster-level effects exist, the between effect might be detected as significant even if the explicit *a* and *b* pathways are 0.

Other results of interest

It is important to note that all methods had decreased performance when pathways were lower strength and when sample sizes were smaller. Also, the ICC had a strong effect on model performance of between indirect effects, which is not surprising, as the ICC quantifies the cluster-level effect, signifying the need for a multilevel model.

This was the first study comparing these three methods in 1-1-1 design with random within-cluster mediation pathways, and thus, we were able to add in a simulation condition that explored the covariance of the *a* and *b* pathways (which in turn changed the overall strength of the within indirect effect). The effect of covariance on the performance of the methods was marked. For all three methods, relative bias of the within indirect effect was larger when the covariance of the pathways was negative (i.e., the within indirect effect was depressed; Table 5). This finding was similar to Bauer et al. (2006). In their study, they found that when estimating multilevel models with random slopes, there was bias in the covariance estimates that was greatest when the *a* and *b* pathways were small. Our study results also suggest that bias was higher when sample size was smaller and the between indirect pathways were larger than the within pathways. Other simulation studies (Preacher et al., 2011) found MSEM performed well, and in fact, better than UMM and CMM for 2-1-1 design with random 1-1 slopes. However, since only the mediator and outcome were measured at level-1, they were unable to explore the covariance of the *a* and *b* pathways. Future research with 1-1-1 designs and random slopes should explore the impact from the covariance in more detail, and confirm if these results hold over other data conditions.

A caution for applied researchers

Before determining which method is appropriate for their data, it is important that applied researchers thoroughly understand the concept of between indirect effects and their interpretation in the context of their research questions. Ultimately, the methodology they chose will depend highly on if they hypothesize the between indirect effects exist in their population, and subsequently, if they are interested in explicitly modeling that effect. However, for both UMM and MSEM, the within and between indirect effects can sometimes seem obscure in their interpretation. The ultimate decision regarding between indirect effects can be driven by the recommendations listed in this document but should ideally be based on a theoretical framework that is supported via exploratory analysis of the relationship between the aggregated mean level-1 predictor, mediator, and outcome.

Our recommendations are limited, but our simulation study suggested that when sample size was small, UMM tended to be more powerful in detecting nonzero between indirect effects, but also exhibited more Type I Error when those effects were truly 0. The empirical example echoed those results; UMM showed significance for between indirect effects with smaller level-1 sample size, but MSEM did not. The empirical example also found larger standard errors for between indirect effects estimated with MSEM. Larger standard errors can have implications on Type I Error, power and CI coverage. In the simulation study, standard errors for the between indirect effects were also much larger on average when using MSEM (M(SE) = .20; SD(SE) = .36) compared to UMM (M(SE) = .03; SD(SE) = .03), and MSEM was found to have higher relative bias in the standard error estimation of between indirect than UMM. This is an especially important consideration for applied researchers to weigh when choosing their methodology. If significance is found with UMM when the sample size is small, it is difficult (probably impossible) to determine if it was due to Type I Error or a correct detection of between mediation. Contrarily, if significance is not found in MSEM, it is likely impossible to determine if this is due to over estimation of standard errors. The performance of the delta method test depends on the accuracy of the standard error estimates. It is possible that UMM underestimates the true sampling error of the between mediation estimates, and thus its usefulness is limited even in cases where it seemed to outperform MSEM. However, based on the results of the simulation study, it seems more likely that MSEM inflates the standard error estimates, especially when the between indirect effect is absent and the ICC is higher. It is also important to note that this problem only occurred in relation to the between indirect effect. Both methods were much more equivalent in relation to the within indirect effects and estimation of corresponding standard errors.

Limitations and directions for future research

We note that our study only applies to variables that are continuous with normally distributed random effects. These methods may perform worse in conditions when random effects were not normally distributed. In a past study, Bauer et al. (2006) found that in the same design (1-1-1 with random slopes), CI coverage with MLM was poorer for nonnormal random effects, as the interval became too narrow under these conditions. Future studies should explore the performance of MSEM under conditions of nonnormal random effects.

Like any Monte Carlo simulation study, our simulation conditions were limited to a specific set of variables. For example, we were more interested in how changes in the between indirect effects affected model performance, and thus, chose not to vary the within effects. Future work should explore how performance of the methods change along with different strengths of within pathways. We also looked at only one instance of Type I Error, although calculated indirect effects can equal 0 in a number of ways in multilevel mediation. We specifically explored conditions where both the *a* and *b* pathways were set to 0, but did not explore the error rate when only one of the pathways was 0 (e.g., $a_B = 0$, $b_B = .4$ would also lead to an indirect effect of 0).

It also is important to discuss that we used a normal theory method (i.e., the delta method) when calculating 95% CI estimate of indirect effects. There are arguments in the literature about preferred methods to determine significance in multilevel models; the normal theory method has been shown to be problematic in CMM when the within and between effects are unbalanced (Zhang, Zyphur, & Preacher, 2009). Despite this limitation, there is support for our use of the normal theory method in this context. First, in a similar study with a 2-1-1 design, the delta method was compared to two alternative methods for determining significance and found to perform similarly in regards to CI coverage in small sample sizes (McNeish, 2017). In another study of multilevel mediation using random slopes, an alternative method (i.e., Monte Carlo method) performed comparably to the delta method when the random effects were normally distributed, as they were in our study (Bauer et al., 2006). Despite this support, general knowledge is limited; future work should focus on alternative methods and compare their performance under differing conditions.

Summary

This study is the first to directly compare the performance of CMM, UMM, and MSEM on a 1-1-1 multilevel mediation design with random slopes. Differences in interpretation of each model are explored within the context of an empirical example. One major advantage of the MSEM approach that cannot be ignored is that it can handle many more designs than UMM or CMM, specifically when the mediator and/or outcome are present at a higher level of the data. However, when the predictor, mediator, and outcome are measured at level-1 (i.e., 1-1-1), the researcher has the option of choosing between CMM, UMM, and MSEM. We recommend that decisions regarding the appropriate modeling method to use should be based on theory, sample size, and the strength of the indirect effects of interest. UMM can be a preferred alternative to MSEM when sample sizes are low, but if sample size requirements are met, MSEM is considered as the preferential method, especially when the between and within pathways are unbalanced. Future work should explore model performance when random effects are nonnormally distributed.

Article Information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors would like to thank Dr. Ming-Te Wang for providing access to the dataset used for the empirical example. The authors are also grateful to the associate editor and 2 reviewers, whose thoughtful suggestions and comments truly strengthened the manuscript. Finally, the authors acknowledge the invaluable contributions of the late Dr. Kevin Kim, whose mentorship allowed for the conception of this project. The ideas and opinions expressed herein are those of the authors alone and endorsement by the authors' institutions is not intended and should not be inferred

ORCID

Christina K. Zigler http://orcid.org/0000-0002-8996-7757

References

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*, 1173–1182. doi:10.1037/0022-3514.51.6.1173

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*(2), 142–163. doi:10.1037/1082-989X.11.2.142

Becker, D., Breustedt, W., & Zuber, C. I. (2017). Surpassing simple aggregation: Advanced strategies for analyzing contextual-level outcomes in multilevel models. *Methods, Data, Analysis, (Online)*, 1–31. doi:10.12758/mda.2017.05

Bradley, J. V. (1978). Robustness. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. doi:10.1111/j.2044-8317.1978.tb00581.x

Croon, M. A., & van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods*, *12*(1), 45–57. doi:10.1037/1082-989X.12.1.45

Eccles, J. S., Wigfield, A., & Scheifele, U. (1997). Motivation to succeed. In W. Damon (Ed.), N. Eisenberg (Series Ed.), *Handbook of Child Psychology* (5th ed.). *Vol. 3*. Social, emotional, and personality development (pp. 1017–1095). New York, NY: Wiley.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233–239. doi:10.1111/j.1467-9280.2007.01882.x

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. doi:10.1037/a0032138

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Publications, Inc.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education*, *22*(1), 1–15. doi:10.3102/0162373707299706

Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, *8*, 157–174. doi:10.1207/S15328007SEM0802_1

Hox, J. J., Maas, C. J., & BrinkhuisKenny, D. A. (2014, October). Mediation. In *David A. Kenny's Homepage*. Retrieved December 11, 2014 from http://davidakenny.net/cm/mediate.htm

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, *8*(2), 115–128. doi:10.1037/1082-989X.8.2.115

Krull, J. K., & MacKinnon, D. P. (1999). Mulilevel mediation modeling in group-based intervention studies. *Evaluation Research*, *23*(4), 418–444. doi:10.1177/0193841X9902300404

Krull, J. K., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*(20), 249–277. doi:10.1207/S15327906MBR3602_06

Ludtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthen, B. (2008). The multilevel latent covariate model: A new more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203–229. doi:10.1037/a0012869

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Taylor & Francis Group, LLC.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavioral Research Methods*, *39*(3), 384. doi:10.3758/BF03193007

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128. doi:10.1207/s15327906mbr3901_4

McNeish, D. (2017). Multilevel mediation with small samples: A cautionary note on the multilevel structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 609–625. doi:10.1080/10705511.2017.1280797

Muthén, B., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. *Longitudinal Data Analysis*, *6*, 143–165. doi:10.1201/9781420011579.ch6

Muthén, L. K., & Muthén, B. O. (1998). *Mplus User's Guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Pituch, K. A., Stapleton, L. M., & Kang, J. Y. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research*, *41*(3), 367–400. doi:10.1207/s15327906mbr4103_5

Pituch, K. A., Whittaker, T. A., & Stapleton, L. M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research*, *40*(1), 1–23. doi:10.1207/s15327906mbr4001_1

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*, 77–98. doi:10.1080/19312458.2012.679848

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, *18*(2), 161–182. doi:10.1080/10705511.2011.557329

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. doi:10.1037/a0020141

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Newbury Park, CA: Sage Publications, Inc.

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavioral Research Methods*, *43*(3), 692–700. doi:10.3758/s13428-011-0076-x

Wang, T., & Degol, J. (2014). Staying engaged: Knowledge and research needs in student engagement. *Child Development Perspectives*, *8*(3), 137–143. doi:10.1111/cdep.12073.

Wang, M.-T., Fredricks, J., Ye, F., Hofkens, T., & Schall, J. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learning and Instruction*, *43*, 16–26. doi:10.1016/j.learninstruc.2016.01.008

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models. *Organizational Research Methods*, *12*(4), 695–719. doi:10.1177/10944281083