The Dissertation Committee for Anita Israni Certifies that this is the approved version of the following dissertation:

Bayesian Estimation of a Longitudinal Mediation Model with Three-Level Clustered Data

Committee:

Susan Natasha Beretvas, Supervisor

Matthew Hersh

Keenan Pituch

Gregory Roberts

Tiffany Whittaker

Bayesian Estimation of a Longitudinal Mediation Model with Three-Level Clustered Data

by

Anita Israni, B.A; M.B.A

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2015

Dedication

Dedicated to my wonderful and supportive parents

Radha and Haku Israni

Acknowledgements

I would like to first extend my sincere gratitude to my truly amazing dissertation advisor, Dr. Tasha Beretvas. She has not only been a wonderful advisor but also a supportive mentor throughout my entire doctoral journey. In our usually weekly meetings, she challenged me intellectually and inspired me to become a better statistician and researcher, all while keeping our meetings entertaining. Her optimism, teachings, sense of humor, brilliance, and dedication to the program and her students are truly admirable. I couldn't have asked for a more understanding advisor who has helped me balance life as a graduate student and a mother.

I would also like to thank my other dissertation committee members who provided insightful feedback and different perspectives. Dr. Keenan Pituch introduced me to the world of hierarchical linear modeling, which was the basis of my dissertation. Dr. Tiffany Whittaker gave me insight into how a different framework might be used. Dr. Greg Roberts offered his invaluable insight from an applied perspective. Dr. Matt Hersh offered unique suggestions on improving estimation of the model. I would also like to thank Dr. Barbara Dodd for her teachings and for introducing me to psychometrics.

I would like to thank my parents, Radha and Haku Israni, for their endless support and love. My mother taught me what unconditional love and support truly is. I hope I can replicate her dedication and selflessness that I experienced in my childhood toward raising my own children. My father has always offered his practical advice on how to handle any challenges professionally and has always stood up for us when my siblings or I needed.

I would also like to thank the support that I have had from my siblings, Vivek, Rakhi, and Rishi, along with all the siblings that I have been fortunate to acquire through my marriage or my siblings' marriages. I have also been so lucky to get support from another wonderful set of parents, Shashi and Chandra Prakash, who have always treated me like their own daughter.

I want to thank my superhero sons, Arman and Arjan, for sacrificing their playtime for my studies and taking care of daddy when mommy was trying to complete her studies. After each of you was born, I was sad to leave you to go back to my research, but both of you motivated me to complete the process. You have brought me so much joy and love, and my life would not be complete without you.

And lastly, to my best friend, my companion, the person who makes me a better person every day, my husband, Arun. While completing my doctoral studies was a good and important decision for me, marrying you is by far the best decision that I have ever made. You have always kept me grounded, supported me, motivated me, and loved me. Your ability to easily forgive and forget, and focus on our relationship and family amazes me. Our sons cannot have a better person to lead them. I love you and thank you from the bottom of my heart. I look forward to our many more adventures to come. Whatever accomplishment this dissertation might represent, it is ours rather than mine.

Bayesian Estimation of a Longitudinal Mediation Model with Three-Level Clustered Data

Anita Israni, Ph.D

The University of Texas at Austin, 2015

Supervisor: Susan Natasha Beretvas

Longitudinal modeling allows researchers to capture changes in variables that take time to exert their effects. Furthermore, incorporating mediation into a longitudinal model allows for researchers to test causal inferences about, for example, how an independent variable might affect growth in an outcome variable through growth in a mediating variable. In scenarios in which multiple variables are measured over time, the parallel process model can be used to model the inter-relationships among the measures' trajectories where both processes are modeled to have their own separate but related growth parameters. The hierarchical linear modeling (HLM) framework can be used to model a parallel process model and allows for easy extensions to handle multiple levels and non-hierarchical data, such as cross-classified or multiple membership data structures, in clustered data.

This study assessed a three-level parallel process model couched in the context of longitudinal mediation where treatment was assigned at the cluster level, matching a longitudinal cluster randomized trial design. The treatment's effect on growth in an outcome is modeled as mediated by the growth in a mediating variable at the cluster and individual level, resulting in a cross-level and cluster-level mediated effect. A simulation and real data analysis study were conducted using a fully Bayesian analysis. In the simulation study, the following four factors were manipulated to assess the recovery of the parameters of interest: mediated effect size, random effects variance component values, number of measurement occasions, and number of clusters.

Overall, relative parameter bias and statistical power improved for higher values for each of the four factors. The cross-level mediated effects were less biased and had greater statistical power than the cluster-level mediated effects. For the mediated effects that were truly zero, coverage rates based on the highest posterior density intervals showed mostly acceptable rates for the cross-level mediated effect and when path *b* was zero paired with a non-zero path *a* for the cluster-level effect. For conditions with a true value of zero for the cluster-level mediated effect with a path *a* of zero, the cluster-level coverage rates provided over-coverage. Results are discussed along with clarification of study limitations and suggestions for future research. Recommendations for applied researchers are also noted.

Table of Contents

List of Tables

List of Figures

.

Chapter 1: Introduction

Very few outcomes, including educational, health and psychological outcomes, are affected immediately by other variables. Instead, it frequently takes time for variables to have their effect, and researchers are commonly interested in understanding patterns of change in variables that might lead to change. Using longitudinal modeling, researchers are able to capture changes in variables that take time to exert their effects (Gollob & Reichardt, 1991). In scenarios in which multiple variables are measured over time, the parallel process model can be used to model the inter-relationships among the measures' trajectories (Cheong, MacKinnon, and Khoo, 2003). With the parallel-process model, the two processes can be modeled to have their own separate but related growth parameters.

As an example, the parallel process model could be used with repeated measures of both mediating and outcome variables to represent inter-relationships among the variables' growth parameters. The independent variable (for example, whether someone was assigned to a treatment or control group) could be modeled as influencing growth in the mediator, which, in turn, could be modeled as affecting growth in the outcome. The structural equation modeling (SEM) framework has been used for this type of longitudinal mediation modeling. Cheong (2011) evaluated a two-level parallel process model in a simulation study and replicated a scenario where measurement occasions were nested within students. However, the SEM framework can be limited in some ways. One example is how time is coded when using SEM for longitudinal growth models. SEM is typically used when the measurement occasions are the same for all individuals or in scenarios with low degrees of heterogeneity in the mean times of measurement between individuals. A high degree of heterogeneity in measurement timing may lead to biases in the SEM parameter estimates (Blozis and Cho, 2008). Another example is the complexity resulting from additional levels of clustering of individuals within contexts, such as students being further nested within schools. While possible in SEM, adding more levels may not be very intuitive with the SEM framework for some researchers. More importantly, while recent versions of the very popular SEM software (Mplus' version 7.0) allow estimation of some basic cross-classified random effects models, SEM estimation procedures and canned software are still not very flexible in how they can be used to handle estimation longitudinal models with data in which individuals are clustered within non-hierarchical clusters (like cross-classified or multiple membership data structures).

The multilevel model or hierarchical linear model (HLM) framework has more flexibility for modeling highly variable measurement occasions for datasets involving repeated measures on individuals and for adding higher clustering levels. In addition, the HLM framework can be more easily used to handle non-hierarchical clustered data structures, such as cross-classified or multiple membership data structures. However, the parameterization of the SEM-based parallel process model is yet to be extended to scenarios for individuals clustered within higher level clusters (i.e, for three level data). In addition, the SEM-based model has not yet been adapted to the multilevel modeling framework. The current study proposes a three-level parallel process model that can be used under the HLM framework and that could be easily extended for use with complex multilevel data structures like cross-classified and multiple-membership data structures. The parallel process model that will be introduced here is couched in the context of longitudinal mediation in which the treatment's effect on growth in an outcome is mediated by the growth in a proximal (mediating) variable. More specifically, two inter-related three-level longitudinal models (e.g., repeated measurements nested within students nested within schools for a mediating and distal outcome variable) will be parameterized under the HLM framework and combined into a single model using dummy-coded variables to allow for estimation of relevant covariances between the models' parameters. A dichotomous independent variable (mimicking an intervention versus control variable) will be hypothesized to influence growth in the mediator variable which will, in turn, influence the growth of the outcome. A mediating effect will be modeled to occur at level two and also level three, resulting in a multiple mediation model where the mediation occurs at the individual and cluster level. A real data analysis will be conducted to demonstrate interpretation of the model's parameters. In addition, a simulation study is proposed to evaluate estimation of the three-level parallel process model.

The following sections will briefly summarize longitudinal models for linear growth in a single variable using the latent variable regression model within the SEM and HLM frameworks. Formulation of the single-level and multilevel mediation models will be presented as well as methods used to test the significance of the mediated effect. The SEM and HLM formulation of a two-level parallel process model with two separate but related growth parameters will next be presented. Last, HLM formulation of a three-level parallel process model using the multilevel modeling framework will be presented before describing the proposed study.

Chapter 2: Literature Review

The current study adds to current mediation literature by proposing a three-level longitudinal mediation model where the growth of the mediator is included as a predictor for the growth of the outcome at level two and level three. The following literature review describes relevant research that has been conducted in the fields of longitudinal models, latent variable regression models, mediation models and longitudinal mediation models under the structural equation modeling and hierarchical linear modeling frameworks. The first section provides an overview of relevant longitudinal and latent variable regression models using the structural equation modeling framework. The next section describes single-level and multilevel mediation models. The following section provides an overview of relevant longitudinal and latent variable regression models using the hierarchical linear modeling framework. The final sections describe a two-level and the proposed three-level longitudinal mediation models followed by a statement of purpose detailing the problem of interest.

Longitudinal Models for Linear Growth in One Variable Using SEM

Using Structural Equation Modeling with Longitudinal Data. The latent growth curve model (LGM) is commonly used for analyzing longitudinal data. The LGM falls within the structural equation modeling (SEM) framework. In the simplest linear LGM, intercept and linear growth rate factors are indicated by repeated measures over time operationalized as total or mean scores on the variable of interest. Figure 1 depicts a typical linear LGM depicting growth in the outcome, Y, across four time points.

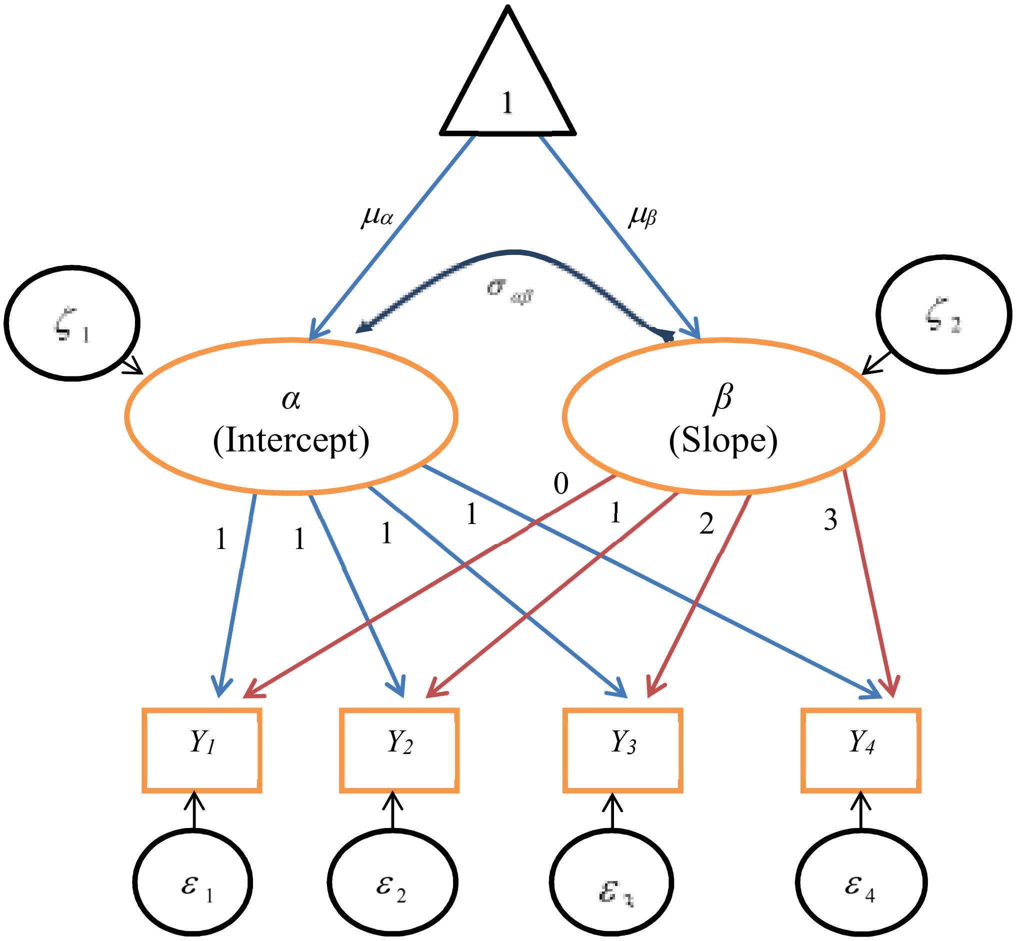

Figure 1. A latent growth model for longitudinal data.

The path coefficients, μ_α and μ_β in Figure 1, represent the overall mean of the intercept and slope factors, respectively. The intercept factor, α, commonly represents the initial value on the outcome variable. This is operationalized by centering the linear growth factor's loadings at the first time point. However, the intercept does not have to reflect the predicted outcome score at the initial measurement occasion. If a researcher intends to assess the predicted *Y* at the final (here, fourth) measurement occasion, then the loadings on the slope factor, β, would be modified from 0, 1, 2 and 3 to −3, −2, −1, and 0, respectively.

If a researcher is interested in modeling linear growth then this is reflected in the coding of the loadings of the indicators on the slope factor, β, in Figure 1. Note, however, that the hypothesized growth rate in the outcome can be modeled assuming any of many functional forms, including, but not limited to, linear, quadratic, logistic, or exponential functions. Researchers may inform their choice of the shape of the growth trajectory through a variety of means including: 1) visually inspecting a plot of the observed data points across time, or 2) setting the factor(s') loadings based on various hypothesized shapes and comparing the associated models' fit, or lastly 3) estimating the growth trajectory's factors' loadings rather than setting them to values. A discussion of these options can be found elsewhere (see, for example, Marsh, Wen, and Hau, 2006); however, the current study will focus solely on linear growth trajectories.

Conventional coding of a linear growth trajectory using the SEM-based LGM typically requires that the linear growth factor's loadings be set to the same values across individuals. Recent research has indicated that rescaling of an age variable either to a common value across all individuals or relative to the first occasion of measurement within each individual would allow estimation of the model with different measurement times for individuals (Mehta and West, 2000). However, the degree of heterogeneity in measurement times may affect parameter estimation in the SEM framework (Blozis & Cho, 2008). A small degree of heterogeneity may allow for the model parameters to be estimated with little or no bias. If respondents vary substantially in the timing of the measurements, then use of the SEM framework for modeling linear growth might not be appropriate and becomes somewhat complex.

The latent means of the intercept and slope factors for the model in Figure 1 provide the predicted initial value on *Y* and the predicted linear growth in *Y*. It is typically assumed that the growth trajectory parameters vary across individuals. Thus, researchers typically estimate the variance in the intercepts and slopes across individuals (represented by the variance terms ζ_α and ζ_β in Figure 1, respectively). Last, it is also frequently assumed (when the data are coded such that the intercept represents the initial measurement occasion) that the intercept covaries with the linear growth. This is captured by the covariance parameter, $\sigma_{\alpha\beta}$, in Figure 1.

One of the benefits of using the SEM framework for analyzing longitudinal data is that directional hypotheses can be assessed. For example, instead of modeling a covariance between the intercept and slope factors, a researcher may wish to model the "effect", γ, of the intercept on linear growth (as depicted in Figure 2).

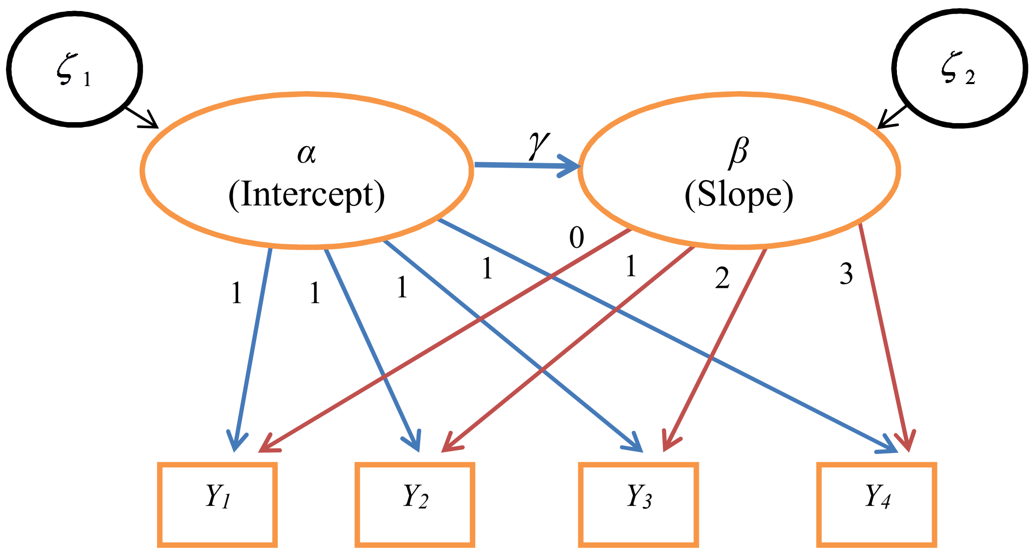

Figure 2. A latent variable regression model for longitudinal data. The error terms for the measured outcomes *(Y_n)* have been omitted to simplify the figure and to focus on the latent variable regression path, γ, which depicts the effect of the intercept on the linear growth.

Use of the SEM framework allows modeling of this kind of hypothesis such that the slope factor, here, would be modeled as an endogenous factor that is predicted by the intercept factor (rather than as an exogenous factor covarying with the intercept factor). Thus, the LGM in Figure 2 involves a latent variable regression model, where one latent variable, the slope, is regressed on another latent variable, the intercept.

The LGM for modeling linear growth in one variable could be extended to handle parallel processes (Cheong et al., 2003). For example, a researcher might be interested in modeling inter-relationships among growth trajectories in two related variables. A specific example of this scenario might involve a longitudinal mediation model in which a researcher might be interested in how growth in a mediator affects growth in an outcome. Before describing how the LGM can be extended to model longitudinal mediation, the next section will briefly summarize the cross-sectional mediation model.

Cross-sectional Mediation Model

Single-level mediation models. The mediation model incorporates a framework for testing hypotheses about chains of causal relationship among multiple variables (MacKinnon, 2008). Figure 3 represents a simple, single-level cross-sectional mediation model in which a primary causal or independent variable, *X*, is modeled as affecting the mediator, *M*, through path *a*, and the mediator is also modeled as affecting the outcome variable, *Y* through path *b*. In some mediation models it is assumed that the mediation is only partial and that in addition to the indirect effect of *X* on *Y* through *M* that there also remains a direct effect of *X* on *Y* (typically represented using *c*'). Much research has been conducted on this simplest cross-sectional mediation model (see, for example, Baron and Kenny, 1986) and on its extensions. The current study is intended to focus on the extension to this mediation model for handling the mediation of a treatment's effect on an outcome's growth by growth in a mediator.

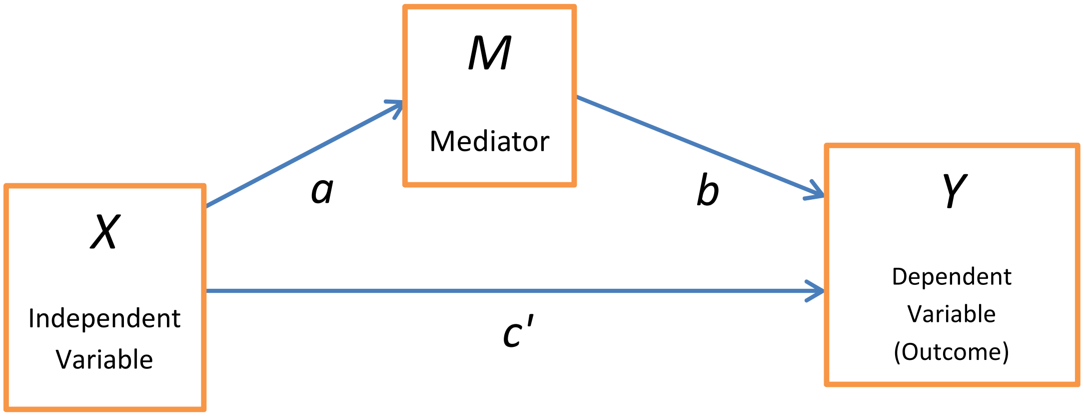

Figure 3. A cross-sectional mediation model shown with the direct and indirect paths.

Given a single-level cross-sectional mediation model where all observations are considered independent, typically a two-step process involving estimation of two regression models is used to estimate the mediation model's parameters. Under the SEM framework, however, estimation of the mediation model's parameters can be accomplished by setting up a single model as depicted in Figure 3. However, as will be shown later in this section, the two-step process can be simplified into a one-step process through the use of dummy-coded variables. This same one-step process for estimating the mediation model parameters will be later extended for use with multilevel data.

If the conventional two-step (single-level) regression model process is used for a single-level mediation model, the first regression equation that must be estimated is as follows:

$$M_i = {}_M\beta_0 + a(X_i) + {}_M\varepsilon_i \tag{1}$$

where M_i is the mediating variable for individual i, ${}_M\beta_0$ is average mediator score controlling for X, and X_i represents the independent variable for individual i (for example, if the individual is in a treatment group or control group). The coefficient for X_i corresponds to path a in Figure 3. The error term for the mediating variable, ${}_M\varepsilon_i$, is typically assumed normally distributed with a mean of zero and a variance of ${}_M\sigma^2$.The following model must then be estimated:

$$Y_i = {}_Y\beta_0 + c'(X_i) + b(M_i) + {}_Y\varepsilon_i \qquad (2)$$

where Y_i is the distal outcome score for individual i and ${}_Y\beta_0$ is the average outcome score controlling for X and M. The coefficients c' and b (see Figure 3) represent the effect of the independent variable on the outcome variable and the effect of the mediating variable on the outcome variable, respectively. The error terms for the dependent variable in Equation 2, ${}_Y\varepsilon_i$, are assumed normally distributed with a mean of zero and a variance of ${}_Y\sigma^2$.

The mediated or indirect effect may be calculated by using one of two procedures. The first procedure is the product of coefficients method that is calculated by multiplying together the two regression coefficients, a and b (Alwin & Hauser, 1975). The second procedure involves calculating the difference between the total effect of the independent variable on the dependent variable without controlling for the mediator, c, and the direct effect of the independent variable on the dependent variable after controlling for the mediator, c' (MacKinnon & Dwyer, 1993). These two estimates of the mediated effect have been shown to be equivalent in an ordinary least-squares regression model (MacKinnon, Warsi, & Dwyer, 1995). However, the current study will focus on the *ab* product estimate of the mediated effect.

Many statistical tests for evaluating the significance of the indirect effects have been proposed and suggested for single-level design data, including Sobel's (1982) test, Baron and Kenny's causal steps (1986), the joint significance test (Cohen & Cohen, 1983), and the asymmetric confidence limits approach (MacKinnon & Lockwood, 2001). Fourteen of these statistical tests, including the ones given above, were compared in a study which used the Type I error rates and statistical power as the basis of comparison (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). In the study presented by MacKinnon et al. (2002), the joint significance test and asymmetric confidence limit approach showed the best balance between the Type I error rates and statistical power for tests of the mediated effect. In the joint significance test, each path of the indirect effect, paths *a* and *b* in the case depicted in Figure 3, is tested for statistical significance. If each path is statistically significant then a mediated effect can be inferred. The asymmetric confidence limits approach provides a confidence interval for the mediated effect around the product of the coefficients, *ab*. The equation for the confidence interval is as follows:

$$Confidence\,Inteval = ab \pm (Critical\,Value_{Upper\,or\,Lower})(\sigma_{ab}) \quad (3)$$

In Equation 3, the parameter, *ab,* represents the meditated effect estimate and σ_{ab} represents the standard error of this effect. The critical values for both the upper and lower confidence intervals is based on a distribution of the product of two normally distributed variables that originated from tables derived by Meeker, Cornwell, and Aroian (1981). While the results for the joint significance test performed slightly better for smaller true effects, use of the asymmetric confidence limits approach provides researchers with confidence interval estimates. The Sobel's (1982) test, which is commonly used in social sciences and is often referred to as the *z* test, calculates a ratio of the estimate of the indirect effect (*ab*) to the Sobel standard error. This ratio is compared to the critical value from a standard normal distribution. However this comparison usually results in low power because it assumes that the *ab* product follows a normal distribution. MacKinnon, Lockwood, and Williams (2004) showed that the product of two normal variates may result in an asymmetric distribution with high kurtosis. As a result, the critical values used with Sobel's test may be too large for the limit of the confidence interval that is closer to zero, resulting in low power. Another commonly used method in research but was also shown to have low power is Baron and Kenny's (1986) approach. This approach proposes a series of tests to assess the significance of an indirect effect. Baron and Kenny proposed three steps to support mediation. Using the paths given in Figure 3, the three steps are as follows: (1) test the significance of the effect of the independent variable on the mediator (path *a*), (2) test the significance of the effect of the mediator on the dependent variable (path *b*), and (3) controlling for the indirect paths (paths *a* and *b*), test the significance of the effect of the independent variable on the dependent variable (path *c'*) to find support for whether the mediation is partial or complete. Before beginning testing each of these casual steps, a researcher must have found an overall statistically significant effect of the independent variable on the dependent variable (path *c*), because the assumption was that if there is no overall effect then mediation cannot occur.

A study by Biesanz, Falk, and Savalei (2010) further evaluated significance tests for indirect effects for single-level design data by including use of non-normal data and introducing some missing data. From the significance tests that were evaluated, a partial posterior predictive distribution method (Bayarri & Berger, 1999, 2000; Robins, van der Vaart, & Ventura, 2000) and the distribution of the product method (MacKinnon, Fairchild, & Fritz, 2007) were shown to have the highest power. The partial posterior predictive distribution method assesses the *p* value for the null hypothesis that no mediation exists. This method further breaks down the null hypothesis into two other null hypothesis, including that Path *a* is equal to zero or Path *b* is equal to zero. Once the *p* value is calculated for each path, the maximum *p* value is used as the *p* value for the test of the mediated effect. The partial posterior method does not provide confidence intervals that are provided by some other methods used for single-level design data, such as with the distribution of the product methods. Biesanz et al. (2010) also evaluated significance tests of indirect effects to assess which tests provided the most accurate and stable coverage rates if a confidence interval was computed. The percentile confidence interval method and the hierarchical Bayesian MCMC approach (Huang Sivaganasen, Succop, & Goodman, 2004) were shown to have the most accurate and stable coverage rates. The hierarchical Bayesian MCMC method determines the posterior distribution of both paths *a* and *b*. Using the posterior distribution of each path, random draws are made and multiplied together to create the confidence interval of the indirect effect, which represents the empirical approximation to the posterior distribution of the indirect effect (*ab*). The percentile confidence interval is computed by a bootstrapping method which takes samples of the same size from the observed data to create an empirical distribution of the indirect effect. This method uses the 2.5% and 97.5% quantiles to represent the 95% confidence interval of the indirect effect. Comparing these two statistical tests, the hierarchical Bayesian MCMC approach performed better in scenarios that had either normal, complete data or non-normal, complete data. Both methods performed equally well in scenarios for normal, incomplete data. The percentile confidence interval performed better in scenarios for non-normal, incomplete data. These results were based on the Serlin's (2000) criterion where decent coverage rates included values falling within the 93.5% to 96.5% range.

As noted earlier, while typically two equations are estimated to test for mediation with single-level data if the SEM framework is not used, a researcher may instead choose to combine these equations through the use of two additional dummy-coded variables. Table 1 depicts a subset of a dataset that has been set up to permit use of this dummy-coding to allow estimation of a single model when estimating the *ab* effect.

Table 1

Data Setup for a One-Step Process Using Dummy-Coded Variables for a Mediation Model.

	Independent Variable	*Dummy Variable for Outcome*	*Dummy Variable for Mediator*	*Outcome Score*	*Mediator Score*	*Outcome for One-Step Equation*
Student	X_i	DY_i	DM_i	Y_i	M_i	Z_i
A	0	1	0	10	8	10
A	0	0	1	10	8	8
B	1	1	0	12	9	12
B	1	0	1	12	9	9
C	0	1	0	14	10	14
C	0	0	1	14	10	10

The columns, DY_i and DM_i in Table 1, are used to identify whether the score contained in variable Z_i represents a score on the mediator or dependent variable for person *i*. The equation to be used for the model estimation using the data setup given in Table 1 is as follows:

$$Z_i = (DM_i)[{}_M\beta_0 + a(X_i) + {}_M\varepsilon_i] + (DY_i)[{}_Y\beta_0 + c'(X_i) + b(M_i) + {}_Y\varepsilon_i] \quad (4)$$

In Equation 4, the outcome variable, Z_i, represents either the mediator or dependent variable based on the values of the dummy-coded variables, DY_i and DM_i. With DY_i equal to one and DM_i equal to zero, the value of Z_i represents the dependent variable. With DY_i equal to zero and DM_i equal to one, the value of Z_i represents the mediating variable. Equation 4 combines Equation 1 and 2 into a single equation. This single equation model, which can be used when the mediator and dependent variables are measured at the same level, cannot be estimated using ordinary least-squares regression due to the pair of residuals, ${}_Y\varepsilon_i$ and ${}_M\varepsilon_i$. However, the model could be estimated using, for example, Bayesian estimation. The single-level mediation equations (Equations 1 and 2 or Equation 4) can be used to estimate the mediated effect, *ab*, when there is no clustering in the dataset. However, if a dataset includes clustered data, such as students within classrooms or schools, then a multilevel mediation model is needed to handle the resulting dependencies. The next section describes multilevel mediation models for cross-sectional data.

Multilevel mediation models for cross-sectional data. Educational research data commonly entails clustering such as is found in datasets with multiple students per classroom or school. In multilevel mediation datasets, the independent variable, mediator, and outcome variables can be measured at any of the relevant levels (for example, at the student or classroom level). Krull and MacKinnon (2001) introduced notation for distinguishing the possible designs differentiated by the level at which each relevant variable is measured. For instance in a 1→1→1 design, the treatment is randomly assigned to individuals (within clusters) and, therefore the independent variable is a student-level variable. In this design, the mediator and dependent variable are each also measured at the individual participant level (level one). This design emulates a multisite randomized control trial. However in a cluster randomized trial, the treatment is administered at the cluster level (level two), but the mediator and outcome might be measured at the student-level. In this case, the design would be labeled a 2→1→1 design (Krull & MacKinnon, 2001).

An example of a two-level 1→1→1 mediation model where the treatment is randomly assigned to students in a multi-site dataset and the mediator and outcome are measured at level one is given below. The baseline level one equation for the 1→1→1 model for the mediator is as follows:

$$M_{ij} = {}_M\beta_{0j} + a\left(X_{ij}\right) + {}_M\varepsilon_{ij} \qquad (5)$$

and at level two:

$${}_M\beta_{0j} = {}_M\gamma_{00} + {}_M u_{0j} \qquad (6)$$

In Equation 5, the effect of the treatment on the mediator is represented by the regression coefficient *a*. In Equation 6, the parameter, ${}_{M}\gamma_{00}$, represents the overall mean value of the mediator's intercept, and ${}_{M}\beta_{0j}$ represents the predicted mean mediator outcome score for classroom *j*. The level one error residuals, ${}_{M}\varepsilon_{ij}$, are typically assumed normally distributed with a mean of 0 and a variance of ${}_{M}\sigma^{2}$. The random effect for classroom *j*'s intercept is ${}_{M}u_{0j}$ and represents the difference between the predicted group-level mean and the observed group-level mean. The level two residuals are also assumed normally distributed with mean of zero and a variance of ${}_{M}\tau_{00}^{2}$. In addition, the regression coefficient, *a*, in Equation 5 is assumed fixed across all level two units; however, a researcher can choose to allow this coefficient to randomly vary across the level two units.

A second set of equations is typically also estimated to provide the measure of the effect of the mediator on the outcome and the effect of the treatment on the outcome (controlling for the other). The baseline equation at level one for this second model is as follows:

$$Y_{ij} = {}_{Y}\beta_{0j} + c'\left(X_{ij}\right) + b\left(M_{ij}\right) + {}_{Y}\varepsilon_{ij} \tag{7}$$

with the following at level-2:

$${}_{Y}\beta_{0j} = {}_{Y}\gamma_{00} + {}_{Y}u_{0j} \tag{8}$$

In the level-one baseline equation given above (Equation 7), the predictor X_{ij} represents the treatment assigned to individual *i* who is a member of classroom *j*, and M_{ij} represents the mediating variable score for student *i* in classroom *j*. In addition, ${}_{Y}\gamma_{00}$ in Equation 8 represents the overall mean intercept, and ${}_{Y}\beta_{0j}$ represents the predicted classroom mean intercept for classroom *j*. The direct effect of the treatment on the outcome is represented by coefficient, *c'*, and the effect of the mediator on the outcome is represented by the regression coefficient, *b*.

These regression coefficients are equivalent to the paths depicted in Figure 3. The level one error residuals, ${}_{Y}\varepsilon_{ij}$, are assumed normally distributed with a mean of 0 and a variance of ${}_{Y}\sigma^{2}$. The random effect for classroom *j*'s intercept is ${}_{Y}u_{0j}$ and represents the difference between the predicted group-level mean and the observed group-level mean. These residuals are also assumed normally distributed with mean of zero and a variance of ${}_{Y}\tau_{00}^{2}$. As with the effect of *a* described in Equations 5, a researcher may choose to have the effect of *b* and *c'* either fixed or randomly varying across the level two units. If both effects (*a* and *b)* are considered randomly varying then the expected value of the indirect (mediated) effect would no longer equal their product. Instead, the covariance between their level two parameters would need be added to the product for an unbiased estimate of the mediated effect (Goodman, 1960).

As with single-level design data, research has also been conducted to assess the performance of statistical tests used for mediated effects for multilevel data. One study by Pituch, Whittaker, and Stapleton (2005) tested the performance of four tests (Sobel, Baron and Kenny, asymmetric confidence limits, and the joint significance approaches) of the mediated effect in the multi-site (two-level) 1→1→1 design in which the intervention was randomly assigned to participants, and participants are members of clusters. Pituch et al. (2005) found that the asymmetric confidence interval estimates exhibited the best power, especially in scenarios with low effect sizes. The authors found that the joint significance test exhibited only slightly less power. Both of these methods also resulted in Type I error rates that were closer to the nominal alpha level as compared with both Baron and Kenny's (1986) and Sobel's (1982) test.

To extend this study, Pituch, Stapleton, and Kang (2006) evaluated some additional tests of the indirect effect for data generated to mimic cluster randomized trials in which clustered groups are randomly assigned to treatment conditions as opposed to randomly assigning individual participants. In Pituch et al.'s (2006) study, three single sample methods were reviewed (the *z* test, empirical-M, and the joint significance test) and three resampling methods were reviewed (the bias-corrected bootstrap, the parametric percentile bootstrap, and the iterated bias-corrected bootstrap). In addition, two versions of cluster randomized models for mediation were investigated, including one version in which the mediator is measured at level one (i.e., the 2→1→1 model) and another version in which the mediator is measured at level two (i.e., the 2→2→1 model). The empirical-M test evaluated in the authors' study is similar to the asymmetric confidence limits test, except that the critical values are based on simulations developed by MacKinnon, Lockwood, and Williams (2004). These new critical values were proposed given that the product may not follow the theoretical distribution used to create the Meeker et al. (1981) tables.

Pituch, Stapleton and Kang (2006) found that the bias-corrected bootstrap test of the indirect effect had the most accurate overall Type I error rates and greatest power. The empirical-M test exhibited the next best performance with power that was only just slightly exceeded by that of the bias-corrected bootstrap. However, both of these tests occasionally exhibited elevated Type I error rates under some scenarios. Therefore, a researcher may want to lower the alpha-level when testing using these methods or use some of the other methods that have lower power but did not elevate the Type I error rates as much. These results were similar across both designs, when the mediator was measured at level one (2→1→1) and when the mediator was measured at the second level (2→2→1).

As shown briefly with the examples presented above, a decent amount of research has been focused on multilevel mediation; however, little research exists investigating mediation from a Bayesian perspective. Yuan and MacKinnon (2009) demonstrated use of a fully Bayesian framework modeling a two-level cross-sectional mediation model with purely hierarchical data. The authors state that the benefits of using Bayesian estimation include the ability to incorporate prior information, the lack of distributional restrictions that are imposed making it appealing for small sample studies, and the potential to fit more complicated, hierarchical models and make inferences about these models in a straightforward and exact way. Yuan and MacKinnon conducted a simulation study to assess parameter recovery for the indirect effect and the covariance between *a* and *b*. Three design factors were manipulated, including the true values of the indirect paths (*a* and *b*), the covariance between the components of the indirect effect, and the level-1 and level-2 sample sizes. The results were compared to results from previous research that had investigated estimation of the same model but using a frequentist estimation procedure. The authors found that the results from the Bayesian approach were comparable to the likelihood-based approach. Negligible bias was found in estimates of the indirect effect and the covariance parameter across conditions. However smaller samples were found to result in more liberal coverage rates. In Yuan and MacKinnon's study, the indirect paths, *a* and *b*, were allowed to vary randomly across level-2 units. The proposed study which will be described in a later section will also evaluate a multilevel mediation model using a fully Bayesian approach. This study extends the work of Yuan and MacKinnon (2009) to include a third level in the model. The proposed model includes two mediation designs including both a 3→3→3 and a 3→2→2 mediation design. The independent variable (treatment) is measured at level three (for example, the school level) emulating a cluster randomized trial with repeated measures (level-1) on the individual (level-2). This independent variable, *X*, will be modeled to have an effect on both the growth of the mediator and the growth of the distal outcome variable. Cluster-level growth in the mediator will also be included as a predictor of growth in the distal outcome at the same level resulting in a 3→3→3 mediation design. In addition, growth in the student's value on the mediator will be modeled as a predictor of growth in the student's value on the distal outcome resulting in a 3→2→2 mediation design. Before describing the proposed model in more detail, the next section will briefly discuss a two-level parallel process model using the SEM framework followed by a discussion of how the models presented here can similarly be parameterized using the HLM framework.

Longitudinal Models for Linear Growth in Two Variables Using SEM

In longitudinal research, researchers might measure multiple variables over time. For example, in a longitudinal intervention study, individuals might be randomly assigned to a treatment or control group and measures of multiple outcomes might be gathered over time. Thus, while the independent variable, X, might be a time-invariant variable assigned at the start of the study, both the mediator and the outcome variables will each have their own growth trajectory. As will be described, the trajectory for both the mediator and the outcome variables could be modeled to have their own growth trajectories with their own hypothesized shapes (for example, linear or quadratic, etc.). In addition, use of the SEM framework would permit modeling inter-relationships among the growth parameters describing the trajectories for the mediator and outcome.

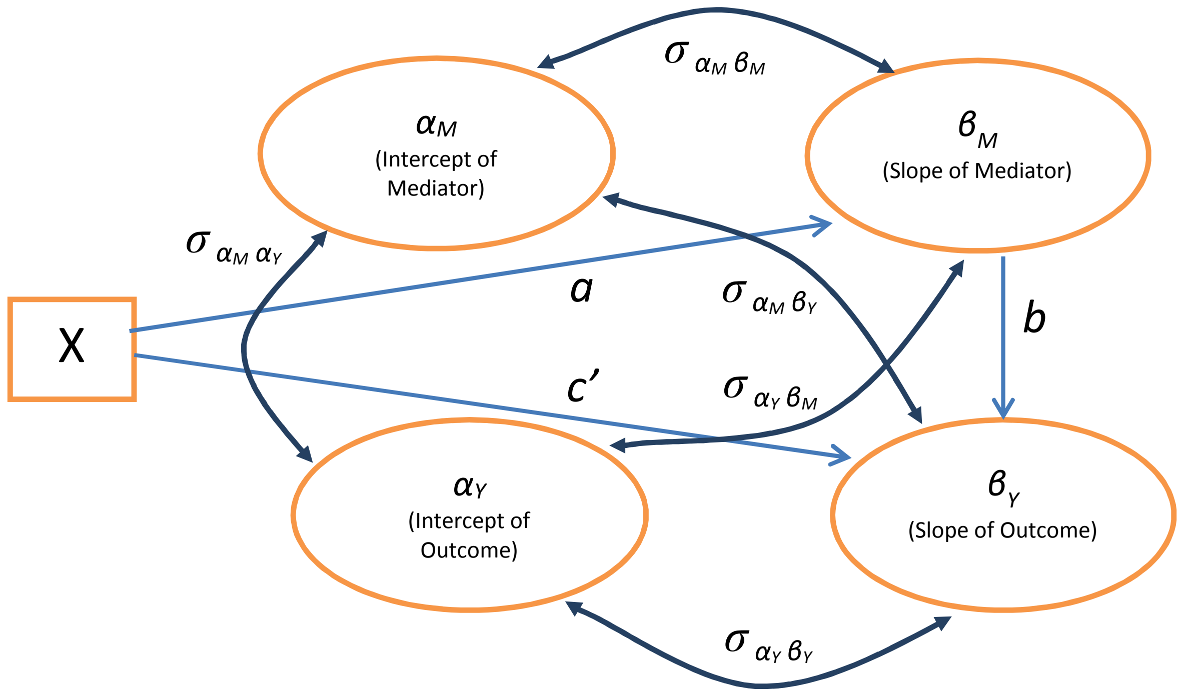

Figure 4. A parallel process latent growth model for mediation. The disturbances, error terms, and observed indicators have been omitted to simplify the Figure and to focus on the direct and indirect effect of the independent variable on the growth of the outcome.

For example, using the simplest linear parallel process LGM, a researcher may hypothesize that a treatment's effect on growth in an outcome is mediated by growth in the mediator (see Figure 4). Matching the notation used in cross-sectional mediation models, Figure 4 includes *a* to represent the effect of the intervention variable, *X*, on linear growth in the mediating variable, *M*, and uses *b* to represent the effect of growth in *M* on growth in the outcome, *Y*. In addition, the direct effect of the intervention on linear growth in *Y* is also represented using c'. Normally the model represented in Figure 4 would also include observed indicators of the intercept, α, and the slope, β, factors for both the mediator, *M*, and outcome variable, *Y*, at each measurement period. However these have been omitted from the Figure to help focus on the parameters of interest capturing the relationships of interest among the intercept and linear growth factors. In the longitudinal mediation model depicted in Figure 4, only the growth rate factor for *M* is modeled as affecting the growth in *Y*. Additional relationships among the latent growth parameters for *M* and for *Y* are modeled as covarying although additional directional relationships could be modeled including, for example, the intercept of the mediator affecting the growth of the outcome variable and/or the intercept of the outcome variable affecting the growth of the outcome variable (see, for example, Cheong, 2011).

Understanding the setup of an SEM model helps elucidate some of the advantages and limitations of using SEM to model longitudinal data. SEM is viewed as a multivariate approach in the sense that each time measurement is represented as a different outcome with its own error term. This setup allows SEM to accommodate complex error and covariance structures (Rovine & Liu, 2012). Another benefit of the SEM framework is its ability to control for measurement error in observed variables. The current study does not focus on this particular benefit of the SEM and leaves that as a direction for future research in the area of longitudinal mediation models. As with any modeling framework, SEM also has its limitations when used to model longitudinal data. Use of the SEM framework for the conventional LGM typically involves constraining the factor loadings for the growth factor to be the same across individuals or rescaling the measurement times to a common value either between or within individuals. Use of SEM is more complicated when individuals are measured at highly varying time points. In addition, while it is possible to use the multilevel SEM (see Preacher, 2011) to handle longitudinal growth models for individuals nested within higher level clustering units (for example, patients within hospitals or students within schools), it can be more complicated to use the SEM rather than the multilevel modeling framework. In addition, it is more complicated to model complex data structures including cross-classified and multiple membership data structures using the SEM framework. The next section presents use of the multilevel modeling framework for modeling growth and extends that model to permit modeling of relationships among parallel process growth parameters.

Use of the Hierarchical Linear Model (HLM) with Longitudinal Data

Use of SEM versus HLM. The SEM and HLM growth curve modeling frameworks have each been used on numerous occasions in applied research to estimate parameters to describe longitudinal data's trajectories. For the most basic single-variable growth trajectory modeling, use of either framework provides corresponding parameter estimates (Stoel, Wittenboer, and Hox, 2003). However, as already mentioned, there are certainly more complicated scenarios in which use of one framework is preferred over use of the other.

One of the distinct advantages of using the HLM over the SEM framework is the flexibility with which measurement occasion timing can be handled. The frameworks have been compared as a multivariate approach (SEM) versus a univariate approach (HLM) (Stoel et al., 2003). With SEM, each of the scores at each measurement occasion is viewed as a dependent variable. And it is typically assumed that each individual is measured at the same occasion; otherwise, the data must be rescaled (Mehta & West, 2000). However, when using the HLM framework for modeling growth, a *Time* variable is included as a predictor at level one. This approach allows for the *Time* variable to take on different values for each individual. This flexibility supports use of the HLM over the SEM framework in situations in which it is difficult to measure individuals at the same or sufficiently similar time periods.

Another advantage for using the HLM over the SEM framework is its flexibility for modeling clustered data with three or more levels. Multilevel educational data are becoming increasing common. Spybrook and Raudenbush (2009) investigated multilevel studies funded by the National Center for Education and the National Center for Education Evaluation and

Regional Assistance and found that 39 studies involved either three or four levels compared to 5 studies that involved only two levels. SEM's LGM framework captures the clustering of measures (level one) within individuals (level two) by modeling measures as multivariate outcomes. And, as noted earlier, it is possible to use the multilevel SEM framework to handle three-level longitudinal data in which, for example, measures are nested within students who might be clustered within schools. However, while possible, multilevel SEM is less commonly used by applied researchers due to the complexity of its estimation and interpretation, and ignoring higher levels may lead to bias parameter estimates. Use of the HLM framework for handling nested data structures may be more intuitive for some researchers and can be more easily extended to handle additional clustering levels and data structure complexities. In addition when, for example, individuals do not attend the same school across time, this complicates the purity of the clustering of individuals within schools through the resulting multiple-membership data structure. Alternatively, students might be cross-classified by neighborhoods and schools. And while it is now possible to use canned SEM software to estimate models for cross-classified data, the options are still quite limited. And it may still be more intuitive both conceptually and operationally for some researchers to use the multilevel modeling framework and software for handling cross-classified or multiple-membership data structures (see, for example, Grady & Beretvas, 2010).

In summary, there are some scenarios in which use of the HLM framework for modeling growth might be preferred by some researchers over the SEM framework. And while one of the primary benefits does involve the flexibility with which more complicated data structures can be handled, the current study is going to focus on the simplest three-level HLM model with a pure hierarchical data structure. Future research can extend what is derived here for the more complex scenarios including clustering of individuals within, for example, cross-classified factors at a higher level. The next few sections describe how the models described earlier using the SEM framework can be parameterized under the HLM framework.

Longitudinal Models for Linear Growth in One Variable Using HLM

The two-level HLM for modeling linear growth in one outcome. Under the HLM framework, modeling growth for longitudinal data is easy to conceptualize when it is recognized that the repeated measures across time points (level one) are nested within individuals (level two). It is necessary to model this dependency either in the SEM framework by explaining correlations between measures across time with a "factor" or by capturing that dependency with a random effect as is done under the HLM framework by recognizing the data as clustered.

The baseline, two-level linear growth model is as follows, at level one:

$$Y_{ij} = \pi_{0j} + \pi_{1j}\left(Time_{ij}\right) + \varepsilon_{ij} \qquad (9)$$

where the dependent variable, Y_{ij}, represents the outcome score of individual j at time i. The time variable, $Time_{ij}$, is a predictor of the dependent variable and represents the time of measurement relative to the baseline measurement period. The parameters, π_{0j} and π_{1j}, represent the individual growth curve parameters where π_{0j} is the baseline measure of individual j at the initial point (due to coding of *Time* with a zero at the first time point) and π_{1j} represents the linear growth rate for that individual. As with the SEM, alternative functional forms can be used and these might impact the coding of the *Time* and polynomial functions of *Time* predictors. However, the current study will focus solely on a linear growth model. The level one error residuals, ε_{ij}, which represent the intra-individual differences across measurement occasions for individual j across time points, are assumed normally distributed with a mean of 0 and a variance of σ^2.

The baseline, two-level linear growth model is as follows, at level two:

$$\begin{cases} \pi_{0j} = \beta_{00} + r_{0j} \\ \pi_{1j} = \beta_{10} + r_{1j} \end{cases} \quad (10)$$

Both growth curve parameters (the intercept and slope) in Equation 10 are decomposed into two parts, the fixed and random effects. The fixed effects, β_{00} and β_{10}, represent the population average initial status and average growth rate, respectively. The random effects, r_{0j} and r_{1j}, represent the inter-individual differences in the initial status and growth rate, respectively. The level two random effects are assumed multivariately normally distributed with means of zero and a covariance matrix of $\begin{bmatrix} \tau^2_{\pi 00} & \tau_{\pi 10} \\ \tau_{\pi 01} & \tau^2_{\pi 11} \end{bmatrix}$.

The three-level HLM for modeling linear growth in one outcome. The two-level HLM longitudinal model can be extended to include three levels if, for example, students are further nested within classrooms or schools, requiring that the dependency within classrooms (or schools) be modeled.

Equations 9 and 10 would be modified to include a subscript *k* to represent the classroom (or school) for individual *j*. The level one equation is as follows:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}\left(Time_{ijk}\right) + \varepsilon_{ijk} \quad (11)$$

with the following level two equations:

$$\begin{cases} \pi_{0jk} = \beta_{00k} + r_{0jk} \\ \pi_{1jk} = \beta_{10k} + r_{1jk} \end{cases} \quad (12)$$

In Equation 11, the dependent variable, Y_{ijk}, now represents the outcome score of individual *j* at time *i* in classroom *k*. In Equation 12, the variables, β_{00k} and β_{10k}, represent the overall intercept and slope for classroom *k,* respectively. The variables, π_{0jk} and π_{1jk} represent the intercept and slope, respectively, for individual *j* in classroom *k*. The random effects, r_{0jk} and r_{1jk}, represent the inter-individual differences for individual *j* in classroom *k* for the intercept and slope, respectively. All other specifications described in Equations 9 and 10 still apply in the three-level model regarding the level one variance and the level two covariance matrix. Extending Equations 11 and 12, the level three equations for the baseline three-level linear growth model is as follows:

$$\begin{cases} \beta_{00k} = \gamma_{000} + u_{00k} \\ \beta_{10k} = \gamma_{100} + u_{10k} \end{cases} \tag{13}$$

As with the level-two growth curve parameters, the level-three growth curve parameters (the intercept and slope) are decomposed into two parts, the fixed and random effects. The fixed effects, γ_{000} and γ_{100}, represent the population average initial status and average growth rate, respectively. The random effects, u_{00k} and u_{10k}, represent the inter-classroom differences in the initial status and growth rate, respectively. The level three random effects are typically assumed multivariately normally distributed with means of zero and a covariance matrix of $\begin{bmatrix} \tau^2_{\beta 00} & \tau_{\beta 10} \\ \tau_{\beta 01} & \tau^2_{\beta 11} \end{bmatrix}$.

Latent Variable Regression Models for Longitudinal Data using HLM

The two-level latent variable regression model using HLM. The SEM framework is more commonly used for estimating a latent variable regression (LVR) model in which, for example, instead of modeling the growth parameters as covarying , the intercept is modeled as influencing the slope. The framework involves two parts in its latent variable regression model, the structural and measurement portion. The latent factors' means and effects are described in the structural portion of the model and are assumed independent of any measurement errors, which are associated with the indicators (measured variables) that are involved in the measurement model. Any error terms associated with the latent factors are considered unsystematic or random terms. The HLM framework also allows for latent variable regression. In its simplest form, the framework formulates a level-one model that describes outcomes as a function of the level-two latent variables. The level-two model further formulates the distribution of the latent variables described by the outcomes. By estimating the joint distribution of the latent variables that vary at two levels, the direct and indirect effects involving these latent variables can be estimated (Raudenbush & Bryk, 2002). For example, the latent intercept parameter could be used as a predictor of the latent rate of change or slope parameter. This allows researchers to assess how a student's growth trajectory is predicted by the initial starting point of the student. Seltzer, Choi, and Thum (2003) demonstrated this use of the HLM framework in their assessment of how math achievement scores change over time based on the initial starting point of students in Grade 7. The authors used the following level two equations:

$$\begin{cases} \pi_{0j} = \beta_{00} + r_{0j} \\ \pi_{1j} = \beta_{10} + \beta_{11}(\pi_{0j}) + r_{1j} \end{cases} \tag{14}$$

along with the level one equation (see Equation 9) to represent their latent variable regression model. Note in Equation 14 that the intercept parameter (π_{0j}) is included as a predictor for the slope parameter (π_{1j}) thereby introducing the latent variable regression model using the HLM framework. The latent variable regression coefficient (β_{11}) represents the amount of change in the growth parameter that is expected to occur given a one unit change in the intercept for person *j*. Additional predictors including interaction terms can, of course, be included to extend these models. However, only the simplest models (that exclude predictors other than *Time* or growth parameters) are discussed here. Future research can be conducted that extends these models with the inclusion of additional predictors at each level.

The assumptions mentioned for the random effects' variances from the model given in Equations 10 still apply; however, the covariance of the random effects is now assumed to be zero in Equations 14 as the growth parameter is conditioned on the intercept. The covariance structure is now represented as $\begin{bmatrix} \tau^2_{\pi 00} & 0 \\ 0 & \tau^2_{\pi 11} \end{bmatrix}$. With the addition of π_{0j} as a predictor of the growth parameter, the slope's residual, r_{1j}, now represents the random effect that exists after controlling for differences in the initial status of students. The authors employed Bayesian estimation using the Gibbs sampler to estimate this model for an applied dataset. HLM software can also be used estimate this type of latent variable regression models in the HLM framework.

The three-level latent variable regression model using HLM. In a later study, Choi and Seltzer (2009) further extended their two-level latent variable regression model using the HLM framework to include a third level to capture the clustering, for example, of students within schools. In this study, in which their three-level latent variable regression model was labeled the LVR-HM3, the authors used Bayesian estimation of the model's parameters. In this three-level design, the level two (student) intercept was included as a predictor of the level two linear growth parameter. The level one equation for Choi and Seltzer's LVR-HM3 model is as follows:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk} a_{ijk} + \varepsilon_{ijk} \tag{15}$$

with the following level two equation:

$$\begin{cases} \pi_{0jk} = \beta_{00k} + r_{0jk} \\ \pi_{1jk} = \beta_{10k} + Bw_k\left(\pi_{0jk} - \beta_{00k}\right) + r_{1jk} \end{cases} \tag{16}$$

Equation 15 is the level one equation for the outcome variable for student *j* in classroom *k* at time *i*. As shown in Equations 16, the intercept at level 2, π_{0jk}, is included as a predictor of the growth parameter, π_{1jk}, at level two. The regression coefficient for the intercept predictor, Bw_k, was labeled by Choi and Seltzer as the within-school (within-site more generically) coefficient. Centering around the group mean of the intercept that is shown in Equation 16 was included to reduce the degree of autocorrelations that can occur in Bayesian estimations. The level three equations of the LVR-HM3 model are as follows:

$$\begin{cases} \beta_{00k} = \gamma_{000} + u_{00k} \\ \beta_{10k} = \gamma_{100} + Bb(\beta_{00k} - \gamma_{000}) + u_{10k} \\ Bw_k = Bw\_0 + Bw\_1(\beta_{00k} - \gamma_{000}) + u_{Bwj} \end{cases} \tag{17}$$

At level three, the intercept of each school, β_{00k}, was also used as a predictor for school *j*'s linear growth parameter, β_{10k}, similar to what was modeled for the student's growth parameter at level two (see Equations 16). The three level model had three outcome parameters for each school: 1) the school mean intercept, β_{00k}, 2) the school mean rate of change, β_{10k}, and 3) the school mean within-school initial status/rate of change slope, Bw_k. For the second and third parameter, the school mean intercept was included as a predictor. When the intercept was used as a predictor for another parameter, the covariances between the random effects of the intercept and that parameter were assumed to equal zero. Therefore, the covariance between the random effects of the school intercept and school growth parameters was set to zero, the covariance between the random effects of the school intercept and school within-school initial status/rate of change slope parameters was set to zero, and the covariance between the random effects of the student intercept and student growth parameters was also set to zero. As a result, the random effects, which are assumed multivariate normally distributed with a vector mean of zero, are now represented with a covariance structure, $\sum$, as follows:

$$\sum = \begin{bmatrix} \tau^2_{\beta 00} & 0 & 0 \\ 0 & \tau^2_{\beta 10} & \tau_{\beta 10, \beta w} \\ 0 & \tau_{\beta w, \beta 10} & \tau^2_{\beta w} \end{bmatrix} \quad (18)$$

The authors (Choi & Seltzer, 2009) conducted a simulation to test estimation of the model's parameters. The authors also demonstrated use of the model by estimating it using real data.

The simulation study examined the effect of using two different priors on the potential bias in the parameters' estimates and on coverage of the interval estimates for the fixed effects parameters. Uniform and inverse gamma (IG) priors were investigated as priors used for the level two variance components. Five fixed effects at level 3 corresponding to those appearing in Equations 14 (γ_{000}, γ_{100}, *Bb*, *Bw_0,* and *Bw_1)* were then evaluated in terms of their degree of bias and the percentage of intervals that included the true value of the corresponding fixed effects parameters. Using uniform priors for the level two variance components resulted in only Bw_0 being slightly negatively biased. The coverage of the 95% credible intervals were all close to 95%. Using IG priors resulted in estimation of Bw_0 being substantially positively biased by 23.2% with poor interval coverage of 65.7%. Use of IG priors also resulted in estimation of Bw_1 being slightly negatively biased. Overall, use of uniform priors proved to result in less relative parameter bias and better interval coverage.

There are scenarios in which a researcher might be interested in inter-relationships among measures of more than one construct over time. For example, a researcher might be interested in a proximal (mediating) variable's trajectory and the relationship between the mediating variable's trajectory parameters and the growth parameters for a distal (outcome) variable. As mentioned earlier, a parallel process model could be used to model changes over time of both the mediator's and outcome's trajectories. The next section describes a two-level HLM-based parallel process model equivalent to the one parameterized using the SEM framework (see Figure 4).

Longitudinal Mediation Models Using HLM

As mentioned above, the parallel process model allows for growth in the mediator and in the outcome to be modeled simultaneously. The most general form of the model allows each process to take its own functional form; however, the current study will focus on the simplest parameterization that assumes both constructs grow linearly over time. The SEM framework is typically used to model parallel processes over time. However, as already noted, while there are benefits to using the SEM framework, there are some limitations to its use and that encourage use of the HLM framework in certain scenarios. The current study is designed to demonstrate how to use the HLM framework to model parallel processes and in particular how to use the framework for longitudinal mediation models ultimately to model the growth in the mediator as a predictor in the growth of the outcome. Before describing the proposed three level longitudinal mediation model, a two-level version of the longitudinal mediation model that regresses the growth of the distal outcome, *Y*, on the growth of the mediator, *M*, will first briefly be summarized.

The two-level longitudinal mediation model using HLM. As noted earlier, a two-level longitudinal mediation model (with measurement occasions nested within individuals) has been parameterized and estimated under the SEM framework (Cheong et al., 2003). For comparison, the same model will be described here using the HLM framework, although no research currently exists that has evaluated estimation of this parameterization of the model. The proposed model, which will be described in a following section, will extend the model described here to three levels. Note that a three-level version of the model has yet to be parameterized under the SEM framework.

In the following section, we will specify a HLM model to have two separate but related growth processes. These models will be combined into one equation using a dummy coded variable to enable the regression of distal outcome trajectory parameters on mediator trajectory parameters. For heuristic purposes, the longitudinal model used for both the mediator and distal outcome assumes linear growth and includes no additional predictors (except for *Time* and other growth model parameters). We specify a parallel process model for the mediator outcome at level 1:

$$M_{ij} = {}_M\pi_{0j} + {}_M\pi_{1j}\left(Time_{ij}\right) + {}_M\varepsilon_{ij} \tag{19}$$

and at level two:

$$\begin{cases} {}_M\pi_{0j} = {}_M\beta_{00} + {}_M r_{0j} \\ {}_M\pi_{1j} = {}_M\beta_{10} + {}_M\beta_{11}(X_j) + {}_M r_{1j} \end{cases} \tag{20}$$

where the dependent variable, M_{ij}, in Equation 19 represents the mediator score of individual *j* at time *i*. The parameters, ${}_M\pi_{0j}$ and ${}_M\pi_{1j}$, represent the individual growth curve parameters for the mediating variable's trajectory where ${}_M\pi_{0j}$ is the baseline measure of individual *j* at the initial point (if *Time* is coded with a zero at the first time point) and ${}_M\pi_{1j}$ represents the linear growth rate for the mediator variable. In Equation 20, the variable, X_j, is a dichotomous variable representing, for instance, whether an individual is in the treatment or control group. The level one error residuals, ${}_M\varepsilon_{ij}$ are assumed normally distributed with a mean of 0 and a variance of ${}_M\sigma^2$. The random effects, ${}_M r_{0j}$ and ${}_M r_{1j}$, represent the inter-individual differences in the initial status and growth rate of the mediator, respectively. The level two random effects are assumed multivariately normally distributed with means of 0 and covariance matrix of $\begin{bmatrix} {}_M\tau^2_{\pi 00} & {}_M\tau_{\pi 10} \\ {}_M\tau_{\pi 01} & {}_M\tau^2_{\pi 11} \end{bmatrix}$.

If interested in modeling growth in the mediator as a predictor of growth in the outcome, then the latent variable regression model similar to the one appearing in Equations 14 could be used where the outcome's linear slope parameter is regressed on the slope for the mediator, ${}_M\pi_{1j}$. The model for the outcome variable would be as follows, at level one:

$$Y_{ij} = {}_Y\pi_{0j} + {}_Y\pi_{1j}\left(Time_{ij}\right) + {}_Y\varepsilon_{ij} \tag{21}$$

and at level two:

$$\begin{cases} {}_Y\pi_{0j} = {}_Y\beta_{00} + {}_Yr_{0j} \\ {}_Y\pi_{1j} = {}_Y\beta_{10} + {}_Y\beta_{11}({}_M\pi_{1j}) + {}_Y\beta_{12}(X_j) + {}_Yr_{1j} \end{cases} \tag{22}$$

As shown in Equation 22, the growth of the mediator (${}_M\pi_{1j}$) for student *j* is included as a predictor of the linear growth parameter of the outcome (${}_Y\pi_{1j}$) for student *j*. Relating the parameterization of this model to the SEM longitudinal mediation diagram, we see that the effect of X_j on growth in *M* is represented by ${}_M\beta_{11}$ in Equations 20 which is equivalent to the *a* parameter in Figure 4. The ${}_Y\beta_{11}$ coefficient in Equations 22 represents the effect of the growth in the mediator, ${}_M\pi_{1j}$, on the growth in the outcome, ${}_Y\pi_{1j}$ which corresponds to coefficient *b* in Figure 4. Last, the coefficient, ${}_Y\beta_{12}$, in Equations 22 representing the direct effect of the treatment on growth in the outcome (controlling for its effect on the mediator's growth) is the equivalent of *c'* in Figure 4. The level two random effects for the outcome are assumed multivariately normally distributed with means of zero and covariance matrix of $\begin{bmatrix} {}_Y\tau^2_{\pi 00} & {}_Y\tau_{\pi 10} \\ {}_Y\tau_{\pi 01} & {}_Y\tau^2_{\pi 11} \end{bmatrix}$.

It is possible and necessary in this latent variable regression model to model both growth trajectories simultaneously. This is possible using dummy coding to identify which outcome (the mediator or outcome) is being modeled. The outcome of the model (Z_{ij}) would be the value for person *j* at time *i* for either the outcome or mediator based on the dummy coded variables' values. The level one equation for the parallel process model using this dummy coding is as follows:

$$Z_{ij} = DY_{ij}[{}_{Y}\pi_{0j} +_{Y} \pi_{1j}(Time_{ij}) +_{Y} \varepsilon_{ij}] + DM_{ij}[{}_{M}\pi_{0j} +_{M} \pi_{1j}(Time_{ij}) +_{M} \varepsilon_{ij}] \quad (23)$$

Equation 23 combines the mediator model given in Equations 19 and 20 and the outcome model given in Equations 21 and 22 into a single model using the dummy-coded variables DY_{ij} and DM_{ij} in the level one equation to identify the relevant outcome. The setup of the corresponding dataset will entail two rows of data for each individual *j* at each measurement occasion *i*. For each row of data, the value for either DY_{ij} or DM_{ij} will be one with the other dummy-coded variable assigned a value of zero. DY_{ij} is the dummy-coded variable identifying that the Z_{ij} value is the outcome score at time *i* for person *j*, and DM_{ij} indicates that the value in Z_{ij} is the mediator score at time *i* for person *j*. Table 2 provides an example of how the data would be set up using the dummy coded variables described and used in Equation 23.

Table 2

Data Setup for a One-Step Process Using Dummy-Coded Variables for a Two-Level Longitudinal Mediation Model with a Treatment Variable.

Student (j)	Treatment (X_j)	Time (i)	DY_{ij}	DM_{ij}	Y_{ij}	M_{ij}	Z_{ij}
A	0	0	1	0	10	8	10
A	0	0	0	1	10	8	8
A	0	1	1	0	12	9	12
A	0	1	0	1	12	9	9
A	0	2	1	0	14	10	14
A	0	2	0	1	14	10	10
B	1	0	1	0	7	6	7
B	1	0	0	1	7	6	6
B	1	1	1	0	9	7	9
B	1	1	0	1	9	7	7
B	1	2	1	0	11	8	11
B	1	2	0	1	11	8	8

The above model provides researchers a parameterization of the two-level longitudinal mediation model using the HLM framework and adds to Cheong's (2011) research, which investigated a similar model using the SEM framework. Cheong conducted a simulation study and manipulated four factors, including the effect size of the mediated effect, explained variances of the measured variables, number of measurement occasions, and sample size. The author found that relative bias of estimates of the mediated effect and its standard error decreased as each one of the factors increased. There was a substantial decrease in relative bias with the addition of two measurement occasions, which the author generated to be three and five. The author also used three methods for testing mediation and empirical power. Given the same conditions, Sobel's first-order solution (Sobel, 1982) had the lowest power compared to the joint significance test (Cohen & Cohen, 1983) or asymmetric confidence interval test (MacKinnon & Lockwood, 2001). The latter two methods were equivalent in power. The ability to model using both frameworks (SEM and HLM) will give researchers more flexibility in choosing the appropriate one for their datasets. Results are expected to be similar for the corresponding factors in the proposed three-level parallel process model.

The two level longitudinal mediation model described above could be used in a research scenario in which a treatment was randomly assigned to individuals (as opposed to being randomly assigned to clusters) (Pituch, Murphy, & Tate, 2010). However, a group or organization may be the focus of an intervention or treatment diffusion may be expected within a cluster, and therefore, implementation of the treatment at the cluster level may be necessary. In this case, a cluster randomized trial design would be utilized. A third level is then needed in the model to appropriately represent the within-cluster dependency.

The following section will now extend the two-level model presented in this section and summarize the proposed three-level longitudinal mediation model using the HLM framework.

The three-level longitudinal mediation model using HLM. A three-level parallel-process model for a dataset consisting of measurement occasions nested within students nested within schools is also a useful model to evaluate. Growth can occur at both level two (for example, for individuals) and three (for example, for schools) and for both the mediator and the distal outcome variables. In such a case, each of the growth parameters (intercept and slope for each outcome) may be modeled to include their own predictors. However the proposed study will focus on the simplest parameterization that includes no additional predictors (except for *Time* and other growth model parameters) and assumes that all growth parameters grow linearly over time.

In the following section, we will specify a multilevel, longitudinal mediation model to handle separate but related growth processes at level two and level three. The same dummy coding idea is applied to enable the regression of distal outcome trajectory parameters on mediator trajectory parameters. We begin by specifying a parallel process model for the mediator outcome at level 1:

$$M_{ijk} = {}_M\pi_{0jk} + {}_M\pi_{1jk}\left(Time_{ijk}\right) + {}_M\varepsilon_{ijk} \tag{24}$$

and at level two:

$$\begin{cases} {}_M\pi_{0jk} = {}_M\beta_{00k} + {}_M r_{0jk} \\ {}_M\pi_{1jk} = {}_M\beta_{10k} + {}_M r_{1jk} \end{cases} \tag{25}$$

Equations 24 and 25 are similar to Equations 19 and 20 described for the two-level version of the model with the exception of the added subscript *k*, which now represents the third level units (schools). The outcome M_{ijk} given in Equation 24 represents the mediator's score for student *j* in school *k* for measurement occasion *i*. At level two, parameter ${}_M\pi_{0jk}$ represents the mediator's intercept for student *j* in school *k* and ${}_M\pi_{1jk}$ represents the mediator's growth for student *j* in school *k*. The overall intercept and growth trajectory for each school *k* are represented by the parameters ${}_M\beta_{00k}$ and ${}_M\beta_{10k}$, respectively. The covariance matrix at level two follow the same specifications described for the two-level longitudinal mediation model in Equations 19 and 20.

In multilevel modeling with longitudinal data, the variance of level-1 error residuals are usually assumed to be constant (homoscedastic) across time measurements as was specified for the multilevel models described above. However, error variances may be a function of a predictor, such as *Time*. For example, if the growth trajectory parameters are expected to explain variances equally across measurement occasions and it is assumed that the proportion of explained variance to total variance remains constant then the error variance would change at each occasion since the level-1 explained variances would increase. This would result in heteroscedasticity of the level one error variances. In some cases, it may be necessary to model this change in error variances. Properly specifying residuals' variances (at any level) is important for inferences about variability in the intercept and slope components and other model parameters' variance components (see, for example, Browne, 2002, Snijders & Bosker, 1999). For the proposed model, the error variances at level one for each measurement occasion will be modeled to have its own unique variance for both the mediator and distal outcome

Extending on Equations 24 and 25, the third level equations representing the schools' growth parameters are as follow:

$$\begin{cases} {}_M\beta_{00k} = {}_M\gamma_{000} + {}_M u_{00k} \\ {}_M\beta_{10k} = {}_M\gamma_{100} + a_B\left(X_k\right) + {}_M u_{10k} \end{cases} \tag{26}$$

In Equation 26, the parameters, ${}_M\beta_{00k}$ and ${}_M\beta_{10k}$, represent the school growth curve parameters for the mediating variable's trajectory where ${}_M\beta_{00k}$ is the baseline measure of school *k* at the initial point (if *Time* is coded with a zero at the first time point) and ${}_M\beta_{10k}$ represents the linear growth rate for the mediator variable. The variable, X_k, is a dichotomous variable representing, for instance, whether a school is in the treatment or control group and replicates a cluster randomized trial, in which groups are randomly assigned to treatments. This design would allow researchers to focus on group interventions and may even prevent treatment diffusion that may dilute the treatment estimates (Pituch, Murphy, & Tate, 2010). The fixed effects, ${}_M\gamma_{000}$ and ${}_M\gamma_{100}$, represent the overall average for the intercept and slope, respectively, for the mediating variable. The regression coefficient, a_B, presents the effect that the treatment has on the growth of the mediator. The random effects, ${}_M u_{00k}$ and ${}_M u_{10k}$, represent the inter-schools differences in the initial status and growth rate of the mediator, respectively. The level three random effects are assumed multivariately normally distributed with means of zero and covariance matrix of

$$\begin{bmatrix} {}_M\tau^2_{\beta 00} & {}_M\tau_{\beta 10} \\ {}_M\tau_{\beta 10} & {}_M\tau^2_{\beta 11} \end{bmatrix}.$$

If interested in modeling growth in the mediator as a predictor of growth in the distal outcome at the student and school level, then the latent variable regression model described in Equations 21 and 22 could be extended to include a third level. The independent variable, *X*, may be retained at level two for a multi-site trial, or *X* may be omitted from level two and included as a predictor in the level-three equations for a cluster-randomized trial. The proposed model will apply to a scenario in which the intervention is randomly assigned at the cluster-level, and therefore, follows a cluster-randomized trial design. The model for the outcome variable would then be as follows, at level one:

$$Y_{ijk} = {}_Y\pi_{0jk} + {}_Y\pi_{1jk}\left(Time_{ijk}\right) + {}_Y\varepsilon_{ijk} \tag{27}$$

and at level two:

$$\begin{cases} {}_Y\pi_{0jk} = {}_Y\beta_{00k} + {}_Y r_{0jk} \\ {}_Y\pi_{1jk} = {}_Y\beta_{10k} + b_{W,k}\left({}_M\pi_{1jk} - {}_M\gamma_{000}\right) + {}_Y r_{1jk} \end{cases} \tag{28}$$

Equations 27 and 28 are similar to the Equations 21 and 22 described for a two-level version of the model with the exception of the added subscript *k*, which now represents the third level unit. The outcome Y_{ijk} given in Equation 27 represents the outcome for student *j* in school *k* at measurement occasion *i*. At level two, parameter ${}_Y\pi_{0jk}$ represents the outcome's intercept for student *j* in school *k* and ${}_Y\pi_{1jk}$ represents the outcome's growth for student *j* in school *k*. The outcome's overall intercept and growth for each school *k* are represented by the parameters

${}_{Y}\beta_{00k}$ and ${}_{Y}\beta_{10k}$, respectively. The growth of the mediator centered around the grand mean of the mediator's growth (${}_{M}\pi_{1jk} - {}_{M}\gamma_{000}$) is included as a predictor of the linear growth parameter of the outcome (${}_{Y}\pi_{1jk}$) for student *j* in school *k*, and its effect is captured by regression coefficient, $b_{w,k}$. The covariance matrix at level two follows the same specifications described for the two-level longitudinal mediation model in Equations 21 and 22. As described above for Equation 24, the level one error variance is commonly assumed to be constant across predictors, such as *Time*, in multilevel modeling. However for the proposed model, the error variance at level one will be assumed to be heteroscedastic.

In Equation 28, the mediator's growth for the individual is centered around its grand mean and included as a predictor of the distal outcome's growth. Various centering options (e.g. group mean or grand mean) may be used to address certain research questions or even help with estimation. For instance in Equations 15 to 17, Choi and Seltzer (2009) incorporated centering into their latent variable regression model to reduce autocorrelation among the Markov chain samples that may occur in Bayesian estimation. Specifically, the authors used group-mean centering in the level-two equation (Equation 16). In mediation models, the use of group-mean centering may pose two problems. First and more importantly, the effect of the treatment on the mediator (a_B) may result in a value of zero (Pituch et al., 2010). With group-mean centering, all groups will have the same mean of zero for the group-centered mediator. If no other predictors are included, the effect of the treatment on the mediator will result in a value of zero, resulting in inaccurate estimates. Second when group-mean centering is used at level two, the cluster-level mediation effect at level three captures the effect of both *b* paths, b_B and b_W (Enders & Tofighi, 2007). To capture the unique (e.g. contextual) effect of the cluster-level mediator on the outcome, grand-mean centering or raw scores should be used. For the proposed model, grand-mean centering will be used.

Building on Equations 27 and 28 for the outcome model, the equations to model the schools' growth parameters would be as follows:

$$\begin{cases} {}_Y\beta_{00k} = {}_Y\gamma_{000} + {}_Y u_{00k} \\ {}_Y\beta_{10k} = {}_Y\gamma_{100} + c'_B\left(X_k\right) + b_B\left({}_M\beta_{10k} - {}_M\gamma_{000}\right) + {}_Y u_{10k} \\ b_{W,k} = b_W \end{cases} \tag{29}$$

In Equation 29, the parameters, ${}_Y\beta_{00k}$ and ${}_Y\beta_{10k}$, represent the school growth curve parameters for the outcome variable's trajectory where ${}_Y\beta_{00k}$ is the baseline measure of school *k* at the initial point (if *Time* is coded with a zero at the first time point) and ${}_Y\beta_{10k}$ represents the linear growth rate for the outcome variable. The variable, X_k, is a dichotomous variable, as described in Equation 26, representing, for instance, whether a school is in the treatment or control group. The third level variable, b_W, represents the within-school effect of the mediating variable's trajectory on the outcome variable's trajectory at level two and is assumed to be fixed and constant across level three. However, a researcher may also choose to allow this effect to vary across level-3 units as was modeled in Choi and Seltzer's (2009) latent variable regression model described in Equation 17 at level three. The fixed effects, ${}_Y\gamma_{000}$ and ${}_Y\gamma_{100}$, represent the overall average for the intercept and slope, respectively, for the outcome variable. The random effects, ${}_Y u_{00k}$ and ${}_Y u_{10k}$, represent the inter-schools differences in the initial status and growth rate of the outcome, respectively. The level three random effects are assumed multivariately normally distributed with means of zero and covariance matrix of $\begin{bmatrix} {}_Y\tau^2_{\beta 00} & {}_Y\tau_{\beta 10} \\ {}_Y\tau_{\beta 01} & {}_Y\tau^2_{\beta 11} \end{bmatrix}$.

As shown in Equation 29, the growth of the mediator ($_{M}\beta_{10k}$) for school k centered around the grand mean of the mediator's growth is included as a predictor of the linear growth parameter of the outcome ($_{Y}\beta_{10k}$) for school k. The effect of X_k on classroom growth in M is represented by a_B in Equations 25 which is equivalent to the a parameter in a mediation model given in Figure 3. The b_B coefficient in Equations 29 represents the effect of the growth in the mediator, $_{M}\beta_{10k}$, on the growth in the outcome, $_{Y}\beta_{10k}$, which corresponds to coefficient b in Figure 3. Last, the coefficient, c'_B, representing the direct effect of the treatment on growth in the outcome (controlling for its effect on the mediator's growth) is the equivalent of *c'* in Figure 3.

The proposed model includes two mediating variables (at level two and level three) and is, therefore, considered a multiple mediator model. Allowing mediation processes to occur at both level two and level three may be necessary if, for example, a researcher hypothesizes that each level measures unique constructs (Pituch & Stapleton, 2012). While research is limited in multiple mediation, literature is available for discussion (see Preacher & Hayes, 2008; MacKinnon, 2000). Following the notation by Krull and MacKinnon (2001), the proposed model follows two designs, a 3→3→3 and a 3→2→2 mediation design. The independent variable is a level three variable and is modeled to have an effect on both the growth of the mediator and the growth of the distal outcome variable at the same level. Cluster-level growth in the mediator is also included as a predictor of growth in the distal outcome resulting in a 3→3→3 mediation design. This design may also be referred to as a cluster-level mediation process, as all paths for the indirect effect exist at the cluster level.

In addition, the independent variable, *X*, is hypothesized to impact growth in the student's value on the mediator which is modeled as a predictor of growth in the student's value on the distal outcome resulting in a 3→2→2 mediation design. This design may also be referred to as a cross-level mediation process, as the treatment occurs at the cluster level and the effect of the mediator on the distal outcome is modeled at the participant level. In multiple mediation model, the total indirect effect becomes the sum of the individual indirect effects, and the total effect of the independent variable on the dependent variable becomes the sum of the direct effect and all individual indirect effects.

As described earlier, use of dummy coding makes it possible to model both growth trajectories simultaneously and estimate a covariance matrix for parameters at each level. The values of the dummy coded variables will identify whether the distal outcome or mediator variable is being modeled. The outcome of the model (Z_{ijk}) would be the value for person *j* in classroom *k* at time *i* for either the outcome or mediator based on the dummy coded variables' values. The level one equation for the parallel process model using this dummy coding is as follows:

$$Z_{ijk} = DY_{ijk}\left[{}_{Y}\pi_{0jk} + {}_{Y}\pi_{1jk}(Time_{ijk}) + {}_{Y}\varepsilon_{ijk}\right] + DM_{ijk}\left[{}_{M}\pi_{0jk} + {}_{M}\pi_{1jk}(Time_{ijk}) + {}_{M}\varepsilon_{ijk}\right] \quad (30)$$

with the following level two equations:

$$\begin{cases} {}_{M}\pi_{0jk} = {}_{M}\beta_{00k} + {}_{M}r_{0jk} \\ {}_{M}\pi_{1jk} = {}_{M}\beta_{10k} + {}_{M}r_{1jk} \\ {}_{Y}\pi_{0jk} = {}_{Y}\beta_{00k} + {}_{Y}r_{0jk} \\ {}_{Y}\pi_{0jk} = {}_{Y}\beta_{10k} + b_{W,k}\left({}_{M}\pi_{1jk} - {}_{M}\gamma_{000}\right) + {}_{Y}r_{1jk} \end{cases} \quad (31)$$

and the level three equations:

$$
\begin{cases}
{}_M\beta_{00k} = {}_M\gamma_{000} + {}_M u_{00k} \\
{}_M\beta_{10k} = {}_M\gamma_{100} + a_B\left(X_k\right) + {}_M u_{10k} \\
{}_Y\beta_{00k} = {}_Y\gamma_{000} + {}_Y u_{00k} \\
{}_Y\beta_{10k} = {}_Y\gamma_{100} + c'_B\left(X_k\right) + b_B\left({}_M\beta_{10k} - {}_M\gamma_{000}\right) + {}_Y u_{10k} \\
b_{W,k} = b_W
\end{cases}
\tag{32}
$$

Equations 30 to 32 combines the mediator model given in Equations 24 to 26 and the outcome model given in Equations 27 to 29 into a single model using the dummy-coded variables DY_{ijk} and DM_{ijk} in the level one equation (Equation 30) to identify the relevant outcome. As noted earlier, multiple mediators are included in the proposed model. The indirect effect for the cluster-level mediation process, where all paths are included at level three, is computed by multiplying the regression coefficients, a_B and b_B. The indirect effect for the cross-level mediation process that includes the effect of the level two mediator's growth on a level two distal outcome's growth is computed by multiplying the regression coefficients, a_B and b_W.

The setup of the corresponding dataset for a three-level longitudinal mediation model is shown in Table 3 and is similar to the setup shown in Table 2 with the exception of an added column to represent the classroom for each individual. The simulated data will follow a purely hierarchical data structure, and therefore, the classroom will be the same for all measurement occasions for each student. However as noted earlier, HLM does have the flexibility to handle multiple membership, where a student may change classrooms during the study's timeframe, and therefore, classroom data for an individual may change across measurement occasions. Further extensions to the proposed model can evaluate estimation with non-hierarchical data structures.

Table 3

Data Setup for a One-Step Process Using Dummy-Coded Variables for a Three-Level Longitudinal Mediation Model with a Treatment Variable.

Student (j)	School (k)	Treatment (X_k)	Time (i)	DY_{ijk}	DM_{ijk}	Y_{ijk}	M_{ijk}	Z_{ijk}
A	1	0	0	1	0	10	8	10
A	1	0	0	0	1	10	8	8
A	1	0	1	1	0	12	9	12
A	1	0	1	0	1	12	9	9
A	1	0	2	1	0	14	10	14
A	1	0	2	0	1	14	10	10
B	1	0	0	1	0	5	7	5
B	1	0	0	0	1	5	7	7
B	1	0	1	1	0	6	7	6
B	1	0	1	0	1	6	7	7
B	1	0	2	1	0	9	11	9
B	1	0	2	0	1	9	11	11
C	2	1	0	1	0	7	6	7
C	2	1	0	0	1	7	6	6
C	2	1	1	1	0	9	7	9
C	2	1	1	0	1	9	7	7
C	2	1	2	1	0	11	8	11
C	2	1	2	0	1	11	8	8

Using the above data setup and the dummy-coded equation described in Equation 30, a Bayesian approach can be used to estimate both models' parameters simultaneously. This specification will also include a covariance matrix at each level (level two and level three) for the relevant model parameters. The level two random effects follow a multivariate normal distribution with means of zero and a 4*x*4 unstructured covariance, $\mathbf{\Sigma}_{\boldsymbol{W}}$ where:

$$\mathbf{\Sigma}_W = \begin{bmatrix} {}_M\tau^2_{\pi 00} & & & \\ \tau_{M\pi 00,M\pi 11} & {}_M\tau^2_{\pi 11} & & \\ \tau_{M\pi 00,Y\pi 00} & \tau_{M\pi 11,Y\pi 00} & {}_Y\tau^2_{\pi 00} & \\ \tau_{M\pi 00,Y\pi 11} & 0 & \tau_{Y\pi 00,Y\pi 11} & {}_Y\tau^2_{\pi 11} \end{bmatrix} \quad (33)$$

It should be noted that in Equation 33 the covariance, $\tau_{M\pi11,Y\pi11}$, between the mediator's growth and distal outcome's growth at level two is now assumed to be zero as the distal outcome's growth is conditioned on the mediator's growth (see Equation 32). The residual for the distal outcome's growth parameter, ${}_Y u_{10k}$, now represents the random effect that exists after controlling for differences in the mediator's growth at the same level. At level three, the 4*x*4 unstructured covariance, $\mathbf{\Sigma}_B$, is as follows:

$$\mathbf{\Sigma}_B = \begin{bmatrix} {}_M\tau^2_{\beta00} & & & \\ \tau_{M\beta00,M\beta11} & {}_M\tau^2_{\beta11} & & \\ \tau_{M\beta00,Y\beta00} & \tau_{M\beta11,Y\beta00} & {}_Y\tau^2_{\beta00} & \\ \tau_{M\beta00,Y\beta11} & 0 & \tau_{Y\pi00,Y\beta11} & {}_Y\tau^2_{\beta11} \end{bmatrix} \qquad (34)$$

where the covariance between the mediator's growth and the distal outcome's growth at level three is assumed to be zero due to the outcome's growth now being conditioned on the mediator's growth. All level three random effects are assumed independent of the level two random effects and level one residuals, and all level two random effects are assumed independent of the level one residuals.

This three-level parallel process model described in Equations 30 to 32 extends upon Cheong's (2011) research investigating a two-level parallel process model. However, this particular model has yet to be applied to real data and its estimation has not been assessed. The current study is intended to demonstrate estimation of this model using a real dataset. In addition, this study will include a simulation study intended to assess parameter recovery under a variety of conditions including the true value of the mediated effect, the proportion of explained variance in the mediator and distal outcome variables, the number of measurement occasions, and the number of clusters.

Statement of Purpose

The proposed study will demonstrate use of the HLM framework for estimation of a three-level parallel process model where two processes can be modeled to have their own separate but related growth parameters occurring at multiple levels. Couched within a longitudinal mediation model, parameter recovery will be investigated when a treatment is modeled as having effect directly and indirectly through the mediator's growth on the distal outcome's growth at the cluster level (group level). In addition, the effect of the growth of the mediator on the growth of the outcome at level two (student level) will also be investigated. Parameterization of the parallel process model within the multilevel modeling framework will not only allow for estimation when use of the SEM framework may be more complicated but will also provide a starting point to evaluate similar models where the hierarchical data structure is more complicated, such as when cross-classified or multiple membership is present. In addition, use of the proposed three-level model allows appropriate handling of data that include clusters of individuals measured over time.

A simulation study is presented that is intended to evaluate estimation of the proposed parallel process model. A number of conditions will be manipulated in the simulation study including the true value of the mediated effect, the proportion of explained variance in the mediator and distal outcome variables, the number of measurement occasions, and the number of clusters.

Parameter recovery for the mediated effects will be assessed for conditions where the mediated effect is a non-zero value by computing the relative parameter bias, empirical power, and empirical coverage rates for the highest posterior density interval estimates. Empirical coverage rates will be computed for conditions in which the true mediated effect is zero through either one or both components of the indirect effect having been generated using a value of zero. Relative parameter bias and empirical coverage rates will also be assessed for the direct effect. Furthermore, an illustrative example will be provided using real data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K).

Chapter 3: Method

A simulation study was conducted assessing estimation of the proposed three-level longitudinal mediation model introduced in the previous section where linear growth in the mediator was modeled as a predictor of linear growth in the distal outcome at both level two (the student level) and level three (cluster or school level). The independent variable, *X*, was simulated to be dichotomous with participation in the intervention randomly assigned at the cluster level (level three) to replicate a cluster randomized trial scenario with repeated measures on individuals. Scores on the mediator and outcome variables were generated to be interval-scaled. Some parameters and design factors for the simulation study were selected based on corresponding research conducted using the SEM framework (Cheong, 2011), as the author investigated a similar two-level longitudinal mediation model. The growth trajectories of both the mediator and the outcome were modeled as linear, and for the sake of convenience (and of correspondence with Cheong's original study), the magnitude of the linear growth for both variables was generated to be the same.

The last measurement occasion for both the mediator and distal outcome variables was generated to be one standard deviation above the average value of the relevant (mediator or outcome) variable at the initial time period for the smaller number of measurement occasions conditions (matching Cheong's study). The resulting slope magnitude computed from the smaller measurement occasion conditions was then used for conditions involving the larger number of measurement occasions to ensure that the same linear trajectory is generated regardless of the length of the study. Even though one of the advantages to using the HLM over the SEM framework is the facility with which highly variable measurement occasions for individuals can be handled, the spacing and number of measurement occasions were simulated to be the same for each individual in this first evaluation of the three-level parallel process model and its estimation using the multilevel modeling framework.

The following equations and covariance structures summarize the values used for the proposed three-level parallel process model. The level three equations for the mediator and distal outcome's growth parameters were generated as follows:

$$\begin{cases} {}_{M}\beta_{00k} = 1.0 + {}_{M}u_{00k} \\ {}_{M}\beta_{10k} = 0.35 + a_B\left(X_k\right) + {}_{M}u_{10k} \\ {}_{Y}\beta_{00k} = 1.0 + {}_{Y}u_{00k} \\ {}_{Y}\beta_{10k} = 0.35 + 0.25\left(X_k\right) + b_B\left({}_{M}\beta_{10k} - 0.35\right) + {}_{Y}u_{10k} \\ b_{W,k} = 0.35 \end{cases} \tag{35}$$

where values for a_B and b_B were generated based on the eight conditions that will be described below and represent the components of the indirect effect for the 3→3→3 mediation design. The regression coefficients, a_B and $b_{W,k}$, shown in Equation 35 represent the components of the indirect effect for the 3→2→2 mediation design.

The level two covariance structure, Σ_W, was generated using the following values:

$$\mathbf{\Sigma}_W = \begin{bmatrix} {}_{M}\tau^2_{\pi 00} & & & \\ \tau_{M\pi 00, M\pi 11} & {}_{M}\tau^2_{\pi 11} & & \\ \tau_{M\pi 00, Y\pi 00} & \tau_{M\pi 11, Y\pi 00} & {}_{Y}\tau^2_{\pi 00} & \\ \tau_{M\pi 00, Y\pi 11} & \tau_{M\pi 11, Y\pi 11} & \tau_{Y\pi 00, Y\pi 11} & {}_{Y}\tau^2_{\pi 11} \end{bmatrix} = \begin{bmatrix} 0.5 & & & \\ -0.1 & 0.1 & & \\ 0.15 & -0.1 & 0.5 & \\ -0.1 & 0 & -0.1 & 0.1 \end{bmatrix} \tag{36}$$

where the covariance of the mediator's growth and distal outcome's growth was generated to be zero because the mediator's growth was included as a predictor for the distal outcome's growth. Values for elements of the level three covariance structure, Σ_B, were generated as follows:

$$\mathbf{\Sigma}_B = \begin{bmatrix} {}_M\tau^2_{\beta 00} & & & \\ \tau_{M\beta 00, M\beta 11} & {}_M\tau^2_{\beta 11} & & \\ \tau_{M\beta 00, Y\beta 00} & \tau_{M\beta 11, Y\beta 00} & {}_Y\tau^2_{\beta 00} & \\ \tau_{M\beta 00, Y\beta 11} & \tau_{M\beta 11, Y\beta 11} & \tau_{Y\beta 00, Y\beta 11} & {}_Y\tau^2_{\beta 11} \end{bmatrix} = \begin{bmatrix} {}_M\tau^2_{\beta 00} & & & \\ -0.1 & {}_M\tau^2_{\beta 11} & & \\ 0.15 & -0.1 & {}_Y\tau^2_{\beta 00} & \\ -0.1 & 0 & -0.1 & {}_Y\tau^2_{\beta 11} \end{bmatrix} \quad (37)$$

where, as with the level two covariance structure, the mediator's growth and the distal outcome's growth at level three in Equation 37 was generated with a covariance of zero, as the distal outcome's growth was conditioned on the mediator's growth.

Following Cheong's (2011) study, the level two intercept variances for the mediator and outcome was 0.5 (for ${}_Y\tau^2_{\pi 00}$ and ${}_M\tau^2_{\pi 00}$, respectively). The variances of the outcome and mediator slope parameters were each generated to have a value of 0.1 (for ${}_Y\tau^2_{\pi 11}$ and ${}_M\tau^2_{\pi 11}$, respectively). In a survey of applied research, L. Muthén and B. Muthén (2002) found an average ratio of about five to one for the ratio of the intercept to the slope variance estimates (i.e., for ${}_Y\tau^2_{\pi 00} / {}_Y\tau^2_{\pi 11}$). The intercepts of the outcome and mediator were simulated to have a correlation of 0.30, mimicking the value used in Cheong's (2011) simulation study. The covariances between the mediator's intercept and outcome variable's slope and between the outcome's intercept and mediator's slope were simulated to be −0.1 (again matching Cheong's values). In addition, the covariances between the intercepts and their corresponding slopes (i.e., for $\tau_{M\pi 00, M\pi 11}$ and $\tau_{Y\pi 00, Y\pi 11}$) were also simulated to be −0.1. This matches research that has indicated that higher starting points lead to smaller rates of change (Seltzer et al., 2003). At level two, the distal outcome's growth was conditioned on the mediator's growth, and therefore, their covariance was assumed to be zero. The distal outcome's trajectory residual at level two, ${}_Y r_{1jk}$, now represents the random effect that exists after controlling for differences in the mediator's growth.

At level three, the true intercept value for both the outcome and mediator (${}_{M}\gamma_{000}$ and ${}_{Y}\gamma_{000}$, respectively) were generated to be 1.0, allowing for the same overall intercept that was used in Cheong's (2011) research.

The level three variance components for the intercept and slope were computed based on previous research, which included an estimated intraclass correlation coefficient, the level one variance, the level two intercept variance from Cheong's (2011) research, and the average ratio of an intercept's variance to its corresponding slope's variance as described earlier from Muthen and B. Muthen's (2002) research. The level one variance was manipulated and will, therefore, result in the level three variance components also being manipulated. Further discussion is given below on the procedure and values used for the level three variance components.

Simulation Study Conditions. The simulation study was conducted using a 6 (true mediated effect size values) x 2 (proportion of variance explained for measured variables) x 2 (number of measurement occasions) x 2 (number of clusters) factorial design resulting from a fully balanced combination of four factors that will be described below. The simulations were carried out using the statistical software package, R, and the model estimation was carried out using Rstan (Version 2.5), which is a Bayesian analysis software package within R. One of the advantages of using the Rstan software package is its ability to set elements within a covariance matrix to certain values and not estimate those elements. This capability was important for this study since the covariance between the growth of the distal outcome and mediator was assumed to be zero, as described in Equations 36 and 37. For each of the 48 conditions, a total of 300 replication datasets were generated and estimated using the model described in Equations 30 to 32.

Computational Resources. The simulations were completed using a high-performance computing system (Stampede) provided by the Texas Advanced Computing Center. This supercomputer allowed for the intensive computation power needed for estimation of the model presented in this study. Stampede runs on a Linux operating system and has 6,400 nodes with 522,080 processing cores (see https://www.tacc.utexas.edu).

To help reduce the estimation completion time, the 3 Markov chains run for each replication were completed in parallel. Using a very small subset of replications, the average estimation completion time for a replication was calculated for different numbers of clusters in the simulated data. The average time per a replication for 20 clusters, 40 clusters, and 100 clusters was 51.95 minutes, 62.69 minutes, and 3.69 hours, respectively.

Number of replications analysis. Analysis was completed for one condition (see the first condition in Table 5) to determine the minimum number of replications from 300, 400, or 500 needed for stable inferences. Little difference was seen across the relative parameter bias of the mediated effects. For instance, the relative parameter bias for the cluster-level effect differed by 0.003 going from 300 to 500 replications. Therefore, 300 replications were simulated for each condition and used for the analysis presented later in this study. In addition given the time required for estimation complete per a replication, 300 replications allowed reasonable completion time of all 48 conditions.

Effect size of the mediated effect. Because recovery of the mediated effect is the focus of this study, the true value of the mediated effects constituted one of the manipulated conditions in this study, mimicking Cheong's (2011) study. The current study investigates various true values for the mediated effects at level two and level three. The mediated effect's effect size is defined as the ratio of the indirect effect to the total effect (which is the sum of the indirect and direct effects). The effect size of the mediated effect at level three for the parallel process model presented in the HLM framework is as follows:

$$\frac{\text{Indirect Effect}}{\text{Total Effect}} = \frac{(a_B)(b_B)}{(a_B)(b_B) + (a_B)(b_W) + c'_B} \tag{38}$$

where a_B, b_B and c'_B represent the *a*, *b* and *c'* paths, as shown in Figure 3, for the indirect effect and direct effect that exists at level three, respectively, and b_W represents the *b* path for the indirect effect at level two. Three non-zero values for the effect size of the mediated effect at level three were generated, namely: 0.08, 0.23, and 0.37 to allow for what would be considered small, medium, and moderately large effect sizes. The values of a_B and b_B from Equation 32, which correspond to the values of *a* and *b*, respectively, in the SEM framework (see Figure 4), were varied to generate the effect sizes and mirror some values examined in Cheong's simulation study (2011). The values for a_B were 0.18, 0.31, and 0.52, and the corresponding values for b_B were 0.16, 0.35, and 0.49. In addition, three zero values were generated for the mediated effect with either a_B or b_B having a zero value paired with a non-zero value for the other parameter, along with one condition in which both a_B and b_B values were generated to be zero. For the zero mediated effect values, in which one parameter is a non-zero value, the parameter with a zero value was paired with the largest value described above for the other (non-zero) parameter. MacKinnon et al. (2002) found that for the methods that performed the best for non-zero mediated effects and for mediated effects when both parameters were generated to be zero were not very accurate when only one path of the mediated effect was generated to be zero. Type I error rates were typically higher when a zero-value parameter was paired with a large effect for the non-zero value. Type I error rates were exceptionally high when path *a* had a value of zero and was paired with a large value for path *b,* compared to path *b* having the zero value and path *a* having a non-zero value. Therefore, conditions with one large effect paired with a zero value were investigated in this study. The values for b_B were generated to be 0 and 0.49, paired with the corresponding values for a_B (which were 0.52 and 0) to represent large values for the non-zero coefficient. The third condition involved datasets generated with true values of zero for both coefficients, a_B and b_B. The direct effect (c'_B), which is equivalent to c' in Figure 4, was generated using a constant value of 0.25 (again matching Cheong's study's generating value). The effect of the mediator's growth on the outcome growth at the student level (b_W) was fixed across level three and generated using a constant value of 0.35. Cheong (2011) found that the medium mediated effect exhibited low relative bias in estimates of the mediated effect for sample sizes of 2,000 and above, and high relative bias for sample sizes of 200 and below. For samples sizes of 500 and 1,000, the medium mediated effect size value exhibited relative bias outside of the acceptable range in only one condition. Using this value allows for the smallest acceptable value for b_W in Equation 32. In addition, using the medium mediated effect size allows investigation in conditions where the *b* path of the indirect effect (b_W) at level two is smaller, equal to, and larger than the *b* path of the indirect effect at level three (b_B). Using a constant generating value of 0.35 for b_W and the three condition-specific values for a_B when the cluster-level indirect effect is generated to be a non-zero value results in the cross-level mediated effect values of 0.06, 0.10, and 0.18, representing small to moderate effect sizes. For the conditions in which a_B, was generated with a value of zero, the true value of the cross-level mediated effect will also be zero.

Explained variances of measured variables. Similar to Cheong's study, heteroscedasticity in the error variances at each measurement time was generated. The value of the error variances was based on a calculated explained variance value and a fixed value for the proportion of explained variance. The equation for the proportion of explained variance that was used to generate the error variance is as follows:

$$\text{Proportion of Explained Variance} = \frac{Explained\ Variance}{Explained\ Variance + Error\ Variance} \quad (39)$$

In Equation 39, the denominator represents the total variance at any given measurement occasion. The explained variance in the outcome (or mediating) variable at each measurement occasion was calculated based on the measurement time values, the variances of the intercept and slope at level two, and their covariances. Using a fixed value for the proportion of explained variance combined with the calculated explained variance at each measurement occasion, the value of the unexplained (error) variance was generated based on Equation 39 for each measurement occasion.

Two proportion values of 0.5 and 0.8 was simulated to replicate a moderate and large percentage of explained variance to total variance, which parallels the values used in Cheong's study. The simulated values were the same for both the mediator and outcome variances for each measurement occasion.

As noted earlier, the level three variance components were generated based on the level one (error) variance, the level two intercept variance, an average intraclass coefficient correlation, and an average ratio of an intercept's variance to its corresponding slope's variance. Heterocedastricity will result in the error variance changing at each measurement occasion to maintain the same proportion of explained variance for the measured variables; therefore for the sake of convenience, the first measurement occasion will be used from each condition (0.5 and 0.8). The level one variance was then assumed constant across the remaining measurement occasions to allow for computation of the level three intercept variance. The level two variances were fixed at values noted above. The average ratio of 5:1 of intercept variance to corresponding slope variance as described earlier from Muthen and B. Muthen's (2002) research (for ${}_{Y}\tau^{2}_{\beta 00}$ and ${}_{Y}\tau^{2}_{\beta 11}$, respectively) were used to compute the slope variance from the resulting intercept variance. The intraclass correlation coefficients (ICC) provides the degree of similarity among units within the same cluster. Hale et al. (2013) investigated ICC values in a three-level longitudinal model, which included measurement occasions nested within students who were then nested within schools, and found that the proportion of unadjusted variance in the outcome at level three (school level) that accounted for academic attainment outcomes ranged from 0.19 to 0.25. Another source to investigate ICC values is the Variance Almanac of Academic Achievement (Hedges & Hedberg, 2007a; Hedges & Hedberg, 2007b), which is a compilation of ICC values and related variance components from a variety of national datasets, including Early Childhood Longitudinal Survey (ECLS) and Longitudinal Study of American Youth (LSAY) data. The values are computed based on various characteristics, including grade level, achievement outcome (mathematics or reading), urbanicity, and socioeconomic status. Selecting characteristics from the Variance Almanac of Academic Achievement similar to the real data variables presented in this study, the unconditional ICC for reading achievement for grade 3 is 0.27, mathematics achievement for grade 5 is 0.21, and reading and mathematics achievement for kindergarten are 0.24 and 0.23, respectively. The authors present ICC values for an unconditional model as well as for various covariate models (e.g. demographics and pretest). The authors showed that ICC values ranged substantially based on characteristics, such as grade level, achievement outcome, and model used. For example, one of the larger ICC values was shown to be 0.32 for reading achievement in grade 9 with one of the smaller values shown to be 0.034 for mathematics achievement in grade 12. For the simulation study, an ICC value of 0.38 was used, mainly to allow for proper estimation of the model, as selecting the ICC value for the current model is one of the variables which determines if the level-three covariance matrix is positive definite as it changes some of the level-three variance values. Using this ICC value resulted in a level three intercept variance generated to be 0.61 with a level three slope variance of 0.12 (for the proportion of explained variance of 0.5 condition), and 0.38 and 0.08 for the intercept and slope, respectively (for the proportion of explained variance of 0.8 condition).

Number of measurement occasions. The number of measurements were simulated to be three and five. The number of measurement occasions were generated to be the same for both the mediator and outcome variable. The slope magnitude of growth parameters at the classroom level were generated using the same value for the mediator and outcome and for each of the two conditions within this factor. As mentioned earlier, the magnitude of the slope for the study were calculated using the three measurement occasions conditions with the outcome at the last measurement occasion generated to be one standard deviation above the average mean of the outcome at the initial time period. With the average intercept value set to 1.0 for both the mediator and the outcome and the corresponding variances of both intercept variables set to 0.5, the condition with three measurement occasions had a final value of 1.707, which is one standard deviation above the initial measurement occasion of 1.0. The magnitude of the slope can then be computed by using the initial and final measurement values. The resulting slope magnitude was 0.35. This slope magnitude was used for both measurement occasion conditions, three and five, and for both variables, the mediator and outcome, to allow for the same linear trajectory to occur if the study were extended to include more measurement occasions. For the five measurement occasions conditions, the last measurement occasion was generated to be 2.4 using the same slope value.

Number of Clusters. The number of clusters at level three was selected based on previous research in cluster randomized trials (Ivers et al., 2011; Pituch, 2005; Spybrook, 2007; Spybrook & Raubenbush, 2009). For example, Spybrook (2007) found that in federally funded cluster randomized trials in education the median number of clusters was 24. Values of 20 and 40 total clusters were used to represent a small and large number of clusters at level three. The number of level two units (students) will be assumed constant at 25. This value is in line with previous research that shows the median cluster size can range from 20 to 34 (Ivers et al., 2011; Spybrook, 2007). Based on these generated values, the total sample size (at level two) was 500 and 1,000 for each set of conditions. Cheong's (2011) research showed that when the level two sample size was less than 500, relative bias was very high for some of the conditions. At a sample size of 500, there were no conditions with unacceptable relative bias for the standard error of the mediated effect and very few with unacceptable relative bias of the mediated effect. Choosing to manipulate the number of clusters discussed above allows for a minimum level two sample size of 500.

Simulation Outcome Measures. The empirical coverage rates, empirical power, and the relative parameter bias for the estimates of the mediated effect at level three (the cluster-level mediation process) and level two (the cross-level mediation process) were assessed. These scenarios include conditions where both indirect paths, *a* and *b* as shown in Figure 3, have non-zero values. The relative bias for the mediated effect estimates were computed as a ratio of difference between the mediated effect estimate and the true value of the effect to the true value of the effect (Hoogland & Boomsma, 1998) as follows:

$$\text{Relativeparameterbias} = RPB(\hat{\theta}) = \frac{\bar{\hat{\theta}} - \theta}{\theta} \tag{40}$$

In Equation 40, both the mean and median value across each condition's 300 replications were assessed for $\overline{\hat{\theta}}$. While the mean is commonly used as an estimate for the parameters of interest, the median may be more robust to a heavy-tailed posterior distribution. Therefore, both values were captured and compared with the analysis. The value of θ represents the true generating parameter's value. The level three mediated effect estimate for the cluster level mediation process were calculated using the product of the equivalent of *a* (equivalent of a_b in Equation 32) and *b* (equivalent of b_b in Equation 32) coefficients as depicted in Figure 3. For the cross-level mediation process, the mediated effect estimate was calculated using the product of the regression coefficients, a_b and b_w from Equation 32. An absolute relative parameter bias of 0.05 or greater will be considered evidence of substantial bias (Hoogland & Boomsna, 1998).

The 95% highest posterior density (*HPD*) intervals were computed to allow assessment of empirical coverage rates and empirical power. The *HPD* intervals are constructed using the posterior distribution and do not require parametric assumptions for their distributional shape. The proportion of replications in which the null hypothesis is rejected (i.e. the interval does not contain zero) was tallied to provide the empirical power of the estimates of non-zero mediated effects. The empirical coverage rates of the estimates of zero and non-zero parameters were computed by tallying the proportion of replications in which the true value of the parameter is included in each *HPD* interval. The criterion for an acceptable coverage rate was determined by first computing the standard error of the nominal coverage probability using the formula as follows (Burton, Altman, Royston, & Holder, 2006):

$$\mathrm{S}E(p) = \sqrt{p(1-p)/B} \tag{41}$$

.

where *SE* represents the standard error, *p* represents the nominal coverage probability, and *B* represents the number of replications used in the analysis. Because 300 replications were used in this study, the value of *SE*(*p*) was 0.126. Burton et al. (2006) suggest that acceptability of coverage should fall within two *SE*s of the nominal rate. Therefore, an acceptable criterion for the coverage rates was defined using a lower limit of 92.48% and an upper limit of 97.52%.

For scenarios in which the mediated effect is truly zero, the Type I error rates of the *HPD* intervals can easily be computed for the cross-level and cluster-level mediation process by subtracting the coverage rate from one. The Type I error rates would then provide the proportion of replications for each condition for which zero is not contained in the *HPD* interval (when the true mediated effect is zero).

Coverage Rate and Relative Parameter Bias for the Direct Effect. *HPD* interval coverage rates were also assessed for the regression coefficient representing the direct effect (c'_B) as shown in Equation 32. Relative bias for the parameter estimate were assessed based on Equation 40. The regression coefficient for the direct effect is a non-zero value of 0.25 that was generated using the same value across all conditions.

Priors used for Bayesian Estimation. Markov chain Monte Carlo (MCMC) estimation was used to estimate the parallel process model proposed. One advantage to using Bayesian estimation is its ability to include prior information in the analysis. Based on the degree to which a researcher wants the data to weigh into the inferences, priors ranging from non-informative to very informative can be used. Non-informative priors may be used if no prior information is available. When non-informative priors are used, estimation of the model will be similar to a model estimated under a frequentist approach. For the first assessment of this model, non-informative priors were used for the parameters of interest, namely the indirect paths, *a* and *b*, and direct path, c', as shown in Figure 3. For these parameters, normal priors were used with means of zero and large variances, which equates to low precision in the Bayesian approach. For the variance components of level 1 and 2 residuals for the proposed multilevel mediation model, non-informative uniform priors were also be used.

Prior distributions for the scalar variance components were selected based on previous research and recommendations. Choi and Seltzer (2009) conducted a simulation study focused on estimation of a three-level latent variable regression model comparing the uniform prior to the inverse gamma prior. The uniform prior was shown to perform better. If the number of clusters is small, then the uniform prior may result in overestimating the variance component cluster-level residuals (Gelman, 2006). The inverse-gamma prior is not recommended due to the inferences becoming sensitive to the prior when a variance component value is close to zero. The half-Cauchy prior is recommended due to its ability to keep the variance values reasonable, its flexibility, and better performance for close-to-zero variance values (Gelman, 2006). Therefore, the half-Cauchy prior was used for all variance components estimated in these models (levels one through three).

The covariance matrix was decomposed into a scale parameter and a correlation matrix, again following recommendations by Gelman (2006) as Stan (the Bayesian estimation software used within Rstan) uses a variant of Hamiltonian Monte Carlo and, therefore, does not have any conjugacy restrictions for multivariate priors. For the correlation matrix, the Lewandowski, Kurowicka, and Joe (lkj) prior was used and paired with the half-Cauchy prior for its scale parameter.

An Illustrative Example Using Data From ECLS-K. To demonstrate the practical application and interpretation of the proposed model within an educational research context, real data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K) was used. The variables selected in this real data analysis were not based on any theory regarding mediational processes underlying achievement outcomes, and therefore, no broader generalizations should be made based on results from this pedagogical analysis. The ECLS-K study followed participants from kindergarten through middle school and gathered a variety of information, including students' kindergarten entry status, school characteristics, and assessment scores from various years. As noted earlier, our simulated model incorporates a dichotomous treatment variable to replicate scenarios where a treatment and control group are being compared. To follow this same type of scenario, schools from the study were separated into two groups based on the type of kindergarten program that the school offered. The "control" group included schools that offered a part-time kindergarten program, and the "treatment" group included schools that offered a full-time kindergarten program. Schools that offered either an AM or PM only program will be considered as part of the "control" group. Schools that offer both were removed from this analysis. Three measurement occasions were used for the mediating variable and outcome variable. Each measurement occasions for the mediator preceded its corresponding measurement occasion for the distal outcome variable to ensure that temporal precedence was established, thus allowing time for the mediator to affect the outcome. In this example the Reading IRT Theta Scores served as the mediating variable, and measurement occasions occurred during the fall 1998 (kindergarten), spring 1999 (kindergarten), and spring 2000 (1st grade). Data from the fall 2000 were omitted due to a high proportion of planned missing values. The Math IRT Theta Scores represented the distal outcome variable, and measurement occasions occurred during the spring 2000 (1st grade), spring 2002 (3rd grade), and spring 2004 (5th Grade). The IRT Theta scores were used instead of the IRT Scale Scores for reasons detailed in Reardon (2007). Only complete cases were included in this illustrative example.

The measurement occasions are spaced out unequally. One of the advantages of the HLM framework is its ability to accommodate uneven intervals in data collection. The *Time* predictor in this dataset corresponded to the semester from the baseline semester that the measurement was taken. The *Time* predictor at the three measurement occasions were 0, 1, and 3 for the mediator and 0, 4, and 8 for the outcome. In addition, the proposed model is intended only for purely hierarchical data and therefore, data for any student who does not attend the same school across all measurement occasions were removed from the dataset that was analyzed. Handling mobile students' data with an extension to this model remains an area for future research. The dataset included 6,070 students who attended the same elementary school during the study's period. There are 724 schools included in the analysis with a mean of 8.38 students per a school and a standard deviation of 4.88. Table 4 provides descriptive statistics for the data set.

Table 4

Descriptive Statistics for the ECLS-K Data

			Part-time Kindergarten (N = 2,773)		Full-time Kindergarten (N = 3,297)	
Outcome	Timing	Variable	*M*	(*SD*)	*M*	(*SD*)
Reading	Fall, Kinder.	M_{0jk}	-1.22	(0.49)	-1.25	(0.51)
	Spring, Kinder.	M_{1jk}	-0.67	(0.46)	-0.66	(0.48)
	Spring, 1st grade	M_{2jk}	0.20	(0.39)	0.16	(0.43)
Math	Spring, 1st grade	Y_{0jk}	0.14	(0.37)	0.09	(0.40)
	Spring, 3rd grade	Y_{1jk}	0.81	(0.35)	0.74	(0.38)
	Spring, 5th grade	Y_{2jk}	1.20	(0.37)	1.12	(0.40)

The three-level longitudinal mediation model described in Equations 30 to 32 with covariance matrices shown in Equations 33 to 34 was fitted to the real data described above. Estimation of regression coefficients, a_B , b_B and b_W , allowed for computation of the cluster-level and cross-level mediated effects. The 95% highest posterior density and credible intervals were computed to estimate a mediated effect at the cluster and individual student levels. In addition, estimation of the regression coefficient, c'_B , and computing its HPD interval allowed for assessment of the direct effect. Analysis of the real data presented here allowed researchers to further understand the practical application of the longitudinal mediation model described within this study in addition to interpretation of the estimates.

Chapter 4: Results

Simulation Study

The purpose of this study was to evaluate parameter recovery when two mediated effects, a cluster-level and cross-level effect, were estimated simultaneously in a three-level longitudinal mediation model, i.e. a parallel process model. The two mediated effects shared a common component, effect a_B, (see Equation 32) which represents the effect of the independent variable on the growth of the mediator at the cluster level. However, each mediated effect has its own b effect, as described in Figure 3. Parameter b_B represents the effect of cluster level's (level-3) growth in the mediator on growth in the distal outcome, and parameter b_W represents the effect of participant level (level-2) growth in the mediator on the corresponding distal outcome's growth. Lastly, a direct effect of the independent variable on the distal outcome's growth at the cluster level was assumed, and is represented by parameter c'_B .

The following chapter details information on convergence and performance across the 48 simulated conditions shown in Table 5, and then results from a real data analysis are also presented. This chapter is divided into five sections. The first section discusses how convergence was assessed and presents convergence rates by grouped conditions. The second section presents and evaluates the relative parameter bias (and parameter bias for truly zero mediated effects) for the parameters of interest including the mediated effects, its components, the fixed effects, and the level-1 variance components. The third section compares coverage rates between the 95% credible and *HPD* intervals of the mediated effects. The fourth section presents the 95% *HPD* interval coverage rates for all parameters of interest. Lastly, a real data set is estimated and analyzed to demonstrate interpretation of parameter estimates.

Table 5

Simulation Study Design Conditions and Generating Parameter Values

Design Conditions		Generating Parameter Values				
					Mediated Effect	
c	m	Prop EV	a_B	b_B	Cross-Level	Cluster-Level
20	3	0.5	0.18	0.16	0.0630	0.0288
			0.31	0.35	0.1085	0.1085
			0.52	0.49	0.1820	0.2548
			0.52	0	0.1820	0.0000
			0	0.49	0.0000	0.0000
			0	0	0.0000	0.0000
		0.8	0.18	0.16	0.0630	0.0288
			0.31	0.35	0.1085	0.1085
			0.52	0.49	0.1820	0.2548
			0.52	0	0.1820	0.0000
			0	0.49	0.0000	0.0000
			0	0	0.0000	0.0000
	5	0.5	0.18	0.16	0.0630	0.0288
			0.31	0.35	0.1085	0.1085
			0.52	0.49	0.1820	0.2548
			0.52	0	0.1820	0.0000
			0	0.49	0.0000	0.0000
			0	0	0.0000	0.0000
		0.8	0.18	0.16	0.0630	0.0288
			0.31	0.35	0.1085	0.1085
			0.52	0.49	0.1820	0.2548
			0.52	0	0.1820	0.0000
			0	0.49	0.0000	0.0000
			0	0	0.0000	0.0000
40	3	0.5	0.18	0.16	0.0630	0.0288
			0.31	0.35	0.1085	0.1085
			0.52	0.49	0.1820	0.2548
			0.52	0	0.1820	0.0000
			0	0.49	0.0000	0.0000
			0	0	0.0000	0.0000
		0.8	0.18	0.16	0.0630	0.0288
			0.31	0.35	0.1085	0.1085
			0.52	0.49	0.1820	0.2548
			0.52	0	0.1820	0.0000
			0	0.49	0.0000	0.0000
			0	0	0.0000	0.0000
	5	0.5	0.18	0.16	0.0630	0.0288
			0.31	0.35	0.1085	0.1085
			0.52	0.49	0.1820	0.2548
			0.52	0	0.1820	0.0000
			0	0.49	0.0000	0.0000
			0	0	0.0000	0.0000
		0.8	0.18	0.16	0.0630	0.0288
			0.31	0.35	0.1085	0.1085
			0.52	0.49	0.1820	0.2548
			0.52	0	0.1820	0.0000
			0	0.49	0.0000	0.0000
			0	0	0.0000	0.0000

Note. Generating value for b_W was 0.35 across all conditions and 25 level-2 units per level-3 unit were generated. c = Number of clusters; m = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance.

Assessing Convergence

Various options are available to assess convergence of MCMC estimated models. Some of those methods include, but are not limited to, visually inspecting trace plots from multiple Markov chains, quantifying the within- and between-chain variance, and monitoring the autocorrelations' plots. In the case of inspecting trace plots for the simulation study, three chains were run in the convergence analysis in a pilot study to assess when mixing of the chains began to occur. At this point of the iteration process, the burn-in period was set to discard the previous samples, and additional iterations were captured from converged chains to ensure suitably accurate inferences. These values were then used for the simulation analysis.

Autocorrelation within a MCMC chain in Bayesian estimation increases the uncertainty of the posterior distribution's estimation. Reduction of autocorrelations can increase mixing of the Markov chains, and therefore, convergence. Thinning is a method used to decrease the correlation between consecutive posterior draws by selecting non-consecutive draws (the n^{th} draw) to compute inferences. In addition, a potential scale reduction measure was reviewed to assess and quantify the ratio of the average variance of samples within each chain to the variance of the pooled samples across chains. As all chains reach equilibrium, the measure converges to one. Gelman and Rubin (1992) recommend that parameters have a measure below 1.1.

Based on the initial convergence analysis, the number of iterations and burn-in samples were set to be the same for all replications and conditions. The number of iterations used for the MCMC estimation was 2,000 with a burn-in of 1,000, using 3 chains, and the thinning value was set to 5. After completion of the estimation of the model using the first set of 300 replication datasets, each parameter estimated within each replication was assessed. If any parameter within a replication did not meet the convergence criterion (i.e. if the scale reduction measure was less than 1.1), then that replication's results were not included in the final analysis and another replication dataset was generated and analyzed.

Table 6 presents the convergence rates across conditions grouped by three of the four factors, which are the number of clusters, number of measurement occasions, and the proportion of explained variance. The fourth factor, mediated effect, had little influence on convergence rates. Conditions with three measurement occasions and a smaller proportion of explained variance (0.5) were shown to be the most problematic, regardless of the number of clusters. The convergence rate for conditions with three measurement occasions and 20 clusters was 19.52% and 22.10% with 40 clusters, which differed substantially from results for the other conditions. The only other condition that was slightly problematic (84.57%) also had 20 clusters and three measurement occasions but was paired with the higher proportion of explained variance (0.8). All other conditions achieved 100% convergence rates. More specifically, all conditions with five measurement occasions showed no problem with convergence. Subsequent replications were simulated for the problematic conditions until all conditions had 300 converged replications.

Table 6
Convergence Rates Across Conditions Grouped by Number of Clusters, Number of Measurement Occasions, and the True Proportion of Explained Level-1 Residuals' Variance

Manipulated Factors			Convergence Rates
c	*m*	Prop EV	Across First Set of 300 Replications[a]
20	3	0.5	19.52%
20	3	0.8	84.57%
20	5	0.5	100.00%
20	5	0.8	100.00%
40	3	0.5	22.10%
40	3	0.8	100.00%
40	5	0.5	100.00%
40	5	0.8	100.00%

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance.
[a]Subsequent replication datasets were generated to replace non-converged replications.

Relative Parameter Bias for Mediated Effects

As discussed earlier, the product of two normal variates, as in the case of both $a_B b_B$ and $a_B b_W$, may result in skewed distribution with high kurtosis (MacKinnon, Lockwood, & Williams, 2004). Therefore, the median of each parameter's posterior distribution was used as the point estimate as it provides a more central value for asymmetric distributions. Table 7 presents the relative parameter bias (*RPB*) for the two mediated effects, cluster-level and cross-level, estimated in the model for scenarios in which the true mediated effect was not zero. The true value of the cluster-level (level-three) mediated effect was calculated as the product $a_B b_B$, and the true value of the cross-level (level-two) mediated effect was calculated as the product $a_B b_W$. While generating values for the a_B and b_B parameters were manipulated by condition, the generating value for b_W remained fixed at 0.35 across all conditions.

As shown in Table 7, all conditions for the cluster-level mediated effect had substantial *RPB* and were negatively biased. The *RPB* values for this mediated effect ranged from -1.006 to -0.105. Overall, increasing the number of clusters decreased the *RPB*, though results were still substantially biased at the higher number of clusters. The average *RPB* decreased by 0.217 going from 20 to 40 clusters. Conditions with 20 clusters had an average *RPB* of -0.592, compared to conditions with 40 clusters with a value of -0.375. However the difference between corresponding conditions with 20 and 40 clusters decreased as the mediated effect increased. For instance, the difference in the average *RPB* for the smallest, non-zero mediated effect was -0.234 going from 20 (average *RPB* = -0.967) to 40 clusters (average *RPB* = -0.733), compared to -0.127 for the largest mediated effect (average *RPB* values of -0.281 and -0.154 for c = 20 versus 40, respectively). In addition, the difference decreased as the number of measurement occasions increased for the larger mediated effect. Comparing conditions with 3 measurement occasions, the difference in average *RPB* was -0.176 for the largest mediated effect going from 20 (average *RPB* = -0.353) to 40 clusters (average *RPB* = -0.177). For conditions with five measurement occasions, the difference -0.077 for the largest mediated effect going from 20 (average *RPB* = -0.209) to 40 clusters (average *RPB* = -0.132). The value of the mediated effect seem to substantially influence the value of the *RPB*. The overall average *RPB* for the smallest, non-zero mediated effect was -0.850, -0.383 for the medium mediated effect, and -0.218 for the largest mediated effect.

The effect of proportion of explained variance on the *RPB* seemed to be dependent on the number of measurement occasions. The average *RPB* decreased going from a proportion of 0.5 (-0.675) to 0.8 (-0.627) for conditions with three measurement occasions and 20 clusters; however, the value increased for conditions with five measurement occasions and 20 clusters going from 0.5 (-0.504) to 0.8 (-0.560). The same pattern occurred for conditions with 40 clusters. The average *RPB* decreased from -0.457 to -0.373 for conditions with three measurement occasions and 40 clusters, and increased from -0.290 to -0.381 for conditions with five measurement occasions and 40 clusters for the two proportion generating values.

Table 7

Relative Parameter Bias for the Non-Zero Mediated Effects' Estimates

Design Conditions		Generating Parameter Values			Mediated Effect	
c	*m*	Prop EV	a_B	b_B	Cross-Level	Cluster-Level
20	3	0.5	0.18	0.16	***-0.193***	***-1.006***
			0.31	0.35	***-0.175***	***-0.654***
			0.52	0.49	***-0.174***	***-0.366***
			0.52	0	-0.004	..
		0.8	0.18	0.16	***-0.055***	***-0.997***
			0.31	0.35	-0.046	***-0.543***
			0.52	0.49	0.005	***-0.340***
			0.52	0	-0.033	..
	5	0.5	0.18	0.16	***-0.218***	***-0.933***
			0.31	0.35	-0.042	***-0.395***
			0.52	0.49	***-0.070***	***-0.186***
			0.52	0	-0.040	..
		0.8	0.18	0.16	***-0.084***	***-0.933***
			0.31	0.35	***-0.069***	***-0.515***
			0.52	0.49	-0.040	***-0.231***
			0.52	0	-0.048	..
40	3	0.5	0.18	0.16	***-0.104***	***-0.921***
			0.31	0.35	***-0.130***	***-0.261***
			0.52	0.49	***-0.089***	***-0.188***
			0.52	0	***-0.065***	..
		0.8	0.18	0.16	-0.026	***-0.697***
			0.31	0.35	-0.011	***-0.258***
			0.52	0.49	0.037	***-0.165***
			0.52	0	-0.043	..
	5	0.5	0.18	0.16	0.038	***-0.570***
			0.31	0.35	-0.008	***-0.195***
			0.52	0.49	-0.008	***-0.105***
			0.52	0	-0.043	..
		0.8	0.18	0.16	***-0.060***	***-0.744***
			0.31	0.35	-0.045	***-0.240***
			0.52	0.49	0.003	***-0.159***
			0.52	0	-0.049	..

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance; .. indicates a true value of zero for the cluster-level mediated effect paired with a non-zero cross-level mediated effect. Bolded, italic values indicate relative parameter bias in excess of the recommended 0.05 cutoff (Hoogland & Boomsma, 1998).

The cross-level mediated effect had a much smaller average *RPB* (-0.059) compared to the cluster-level mediated effect (-0.483). Of the conditions that were substantially biased for the cross-level mediated effect, all were also negatively biased, similarly to the cluster-level mediated effect. However, there were conditions where no substantial *RPB* was found. One case included conditions with a non-zero cross-level mediated effect (0.182) paired with a cluster-level mediated effect of zero. These conditions had a generating value of zero for the b_B parameter but not for the a_B parameter. All but one of these conditions were within the *RPB* cutoff of 0.05, and the average *RPB* for this set of conditions was -0.041.

The *RPB* improved as the number of clusters increased, as was also found for the cluster-level mediated effect. The average *RPB* for conditions with 20 clusters was -0.080 compared to -0.038 for conditions with 40 clusters. Unlike in the cluster-level mediated effect, as the proportion of explained variance increased, the average *RPB* improved except in one set of conditions that had 40 clusters and 5 measurement occasions. For 20 clusters, the average *RPB* was -0.137 and -0.032 for 3 measurement occasions and -0.093 and -0.060 for 5 measurement occasions when the proportion of explained variance was 0.5 and 0.8, respectively. For 40 clusters, the average *RPB* was -0.097 and -0.011 for 3 measurement occasions and -0.005 and -0.038 for 5 measurement occasions when the proportion of explained variance was 0.5 and 0.8, respectively. For conditions with 20 clusters, increasing the value of the mediated effect also improved the *RPB* substantially. The smallest, non-zero mediated effect had an average *RPB* of -0.138, and the largest mediated effect with a non-zero cluster-level effect had an average *RPB* of -0.070. For conditions with 40 clusters, the difference was less substantial. The conditions with a smaller, non-zero mediated effect conditions had an average *RPB* of -0.038 compared to -0.014 for the larger mediated effect with a non-zero cluster-level effect. Conditions with 40 clusters, three measurement occasions, and a proportion of explained variance of 0.5 were also problematic. All conditions within this set had substantial *RPB*. There was only one other condition with 40 clusters that was slightly substantially biased at -0.065. Overall, conditions with 40 clusters recovered the cross-level mediated effect better. Grouped by the number of measurement occasions, the number of clusters, and the proportion of explained variance, conditions with the smallest value in each of these factors (20 clusters, 3 measurement occasions, and 0.5 proportion of variance explained) had the highest average *RPB* at -0.137. The set of the conditions with the lowest average *RPB* (-0.005) included 40 clusters, 5 measurement occasions, and 0.5 proportion of explained variance. Overall, *RPB* improved as the number of clusters, the proportion of explained variance, mediated effect size, and the number of measurement occasions increased.

Parameter Bias for Mediated Effects

Table 8 presents the parameter bias (*PB*) for both mediated effects including scenarios when the true mediated effect was zero and when it was non-zero. When the cluster-level mediated effect was zero, its parameter bias was on average lower (-0.012) compared with conditions with a non-zero cluster-level mediated effect (-0.040). Broken down by number of clusters, the conditions with 20 clusters and a non-zero cluster-level mediated effect had an average *PB* of -0.052 compared to -0.016 for the conditions with a zero cluster-level mediated effect and 20 clusters. Within conditions with 40 clusters, the values were -0.029 versus -0.008 for non-zero and zero value cluster-level mediated effects, respectively. The *PB* increased when the mediated value increased from a small, non-zero value to a larger value. The true value of the cluster-level mediated effect increased 7.847 times in the simulation conditions going from the small to large mediated effect, while the average *PB* increased 1.553 times for the conditions with 20 clusters and 0.869 times for the conditions with 40 clusters.

The cross-level mediated effect also showed a difference in the average *PB* for zero, cross-level mediated effects (-0.0004) compared to non-zero, cross-level mediated effects (-0.007) conditions. The average *PB* for the cross-level mediated effect for conditions with 20 clusters were -0.010, -0.006, -0.0009 for conditions with a non-zero cross-level paired with a non-zero cluster-level mediated effect, a non-zero cross-level paired with a zero cluster-level mediated effect, and both zero mediated effects, respectively. For conditions with 40 clusters, the average *PB* were -0.003, -0.009, and 0.0004.

Relative Parameter Bias for Direct and Indirect Effects

Table 9 reports the *RPB* for the direct effect and each of the individual effects that comprise the mediated effects (a_B, b_B, and b_W). Results indicated that estimates of b_W exhibited substantial, positive *RPB* for five of the six conditions with 20 clusters, 3 measurement occasions, and a 0.5 proportion of explained variance (i.e. the smallest value for each of the factors manipulated). The average *RPB* for this set of conditions is 0.124. All but one other condition (which was biased at -0.061) resulted in acceptable, unsubstantial *RPB* values.

Effect a_B was estimated with substantial *RPB* in three conditions, two of which have a small (non-zero) value for this effect (0.18). The third condition with unacceptable *RPB* is included in the first set of conditions where b_W also had a high number of conditions with unacceptable bias. The problematic effects estimated were b_B and c'_B . Parameter b_B was substantially negatively biased for all conditions. However, most of the conditions resulted in substantial, positive bias for parameter c'_B . The correlations between the parameter bias (*PB*) of these two parameters is -0.360. The negative correlation indicates that to some small extent as one of the parameters increases, the other parameter decreases. Parameter c'_B was over-estimated for many of the conditions, while b_B was underestimated for several corresponding conditions. The number of clusters seems to impact the *RPB* in estimates of b_B with the values of -0.367 and -0.194 for 20 and 40 clusters, respectively. For c'_B, the average *RPB* values were 0.102 and 0.060 from 20 and 40 clusters, respectively. The *RPB* for parameter b_B did improve as its true value increased. For instance the average *RPB* for the conditions with the smallest value of b_B (0.16) was -0.596. For the conditions with the largest value of b_B (0.49 paired with the non-zero effect a_B) was -0.150. However, the value of a_B did not seem to affect the *RPB* for parameter b_B when this effect was large. For instance, the largest value of b_B was also paired with a zero value of a_B. The average *RPB* was -0.154 for this set conditions compared to -0.150 when paired with a non-zero value of a_B. As the proportion of variance explained increased from 0.5 to 0.8, the *RPB* for b_B also increased. Within conditions with 20 clusters, the average *RPB* for the conditions with 0.5 proportion was -0.285 compared to -0.449 for conditions with 0.8 proportion. For 40 clusters, the average *RPB* values were -0.155 and -0.233 for 0.5 and 0.8 proportion conditions, respectively. The average *RPB* also decreased going from 3 (-0.335) to 5 (-0.225) measurement occasions.

The *RPB* for c'_B did not seem to be influenced by whether b_B had a zero or non-zero value. The average *RPB* for parameter c'_B for 20 clusters was 0.096 and 0.113 for non-zero versus zero b_B, and for the 40 clusters, the average *RPB* was 0.062 and 0.056 for non-zero versus zero b_B conditions. However, the number of clusters did seem to influence recovery of c'_B. The average *RPB* for 20 clusters was 0.102 versus 0.060 for 40 clusters. In conditions where there was a cluster-level mediated effect with a generating value of zero paired with a non-zero cross-level mediated effect, the average *RPB* was 0.186.

Table 8

Parameter Bias for the Mediated Effects' Estimates

Design Conditions		Generating Parameter Values			Mediated Effect	
c	*m*	Prop EV	a_B	b_B	Cross-Level	Cluster-Level
20	3	0.5	0.18	0.16	-0.012	-0.029
			0.31	0.35	-0.019	-0.071
			0.52	0.49	-0.032	-0.093
			0.52	0	-0.001	-0.045
			0	0.49	-0.001	0.000
			0	0	-0.001	0.000
		0.8	0.18	0.16	-0.003	-0.029
			0.31	0.35	-0.005	-0.059
			0.52	0.49	0.001	-0.087
			0.52	0	-0.006	-0.072
			0	0.49	0.001	0.000
			0	0	0.004	0.000
	5	0.5	0.18	0.16	-0.014	-0.027
			0.31	0.35	-0.005	-0.043
			0.52	0.49	-0.013	-0.047
			0.52	0	-0.007	-0.020
			0	0.49	-0.003	0.000
			0	0	-0.004	0.000
		0.8	0.18	0.16	-0.005	-0.027
			0.31	0.35	-0.007	-0.056
			0.52	0.49	-0.007	-0.059
			0.52	0	-0.009	-0.059
			0	0.49	-0.002	0.000
			0	0	-0.002	0.000
40	3	0.5	0.18	0.16	-0.007	-0.027
			0.31	0.35	-0.014	-0.028
			0.52	0.49	-0.016	-0.048
			0.52	0	-0.012	-0.034
			0	0.49	0.002	0.001
			0	0	0.000	0.000
		0.8	0.18	0.16	-0.002	-0.020
			0.31	0.35	-0.001	-0.028
			0.52	0.49	0.007	-0.042
			0.52	0	-0.008	-0.034
			0	0.49	-0.002	0.000
			0	0	0.001	0.000
	5	0.5	0.18	0.16	0.002	-0.016
			0.31	0.35	-0.001	-0.021
			0.52	0.49	-0.002	-0.027
			0.52	0	-0.008	-0.006
			0	0.49	0.000	0.000
			0	0	0.000	0.000
		0.8	0.18	0.16	-0.004	-0.021
			0.31	0.35	-0.005	-0.026
			0.52	0.49	0.000	-0.040
			0.52	0	-0.009	-0.017
			0	0.49	0.001	0.000
			0	0	0.000	0.000

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance.

Table 9

Relative Parameter Bias for the Indirect and Direct Effects Estimates

Design Conditions		Generating Parameter Values			Effects			
c	m	Prop EV	a_B	b_B	a_B	b_B	c'_B	b_W
20	3	0.5	0.18	0.16	0.013	***-0.639***	-0.032	***0.112***
			0.31	0.35	-0.024	***-0.351***	***0.084***	***0.170***
			0.52	0.49	***-0.082***	***-0.195***	***0.083***	0.042
			0.52	0	-0.032	..	***0.184***	***0.119***
			0	0.49	..	***-0.269***	***-0.066***	***0.135***
			0	0	..	..	0.026	***0.168***
		0.8	0.18	0.16	0.023	***-1.076***	***0.115***	0.032
			0.31	0.35	-0.006	***-0.429***	***0.180***	0.001
			0.52	0.49	0.001	***-0.279***	***0.237***	0.042
			0.52	0	-0.012	..	***0.313***	-0.011
			0	0.49	..	***-0.283***	-0.016	0.004
			0	0	..	..	-0.028	0.027
	5	0.5	0.18	0.16	-0.037	***-0.537***	***0.083***	0.011
			0.31	0.35	0.021	***-0.077***	0.047	0.010
			0.52	0.49	-0.007	***-0.082***	***0.177***	0.013
			0.52	0	-0.026	..	***0.103***	0.023
			0	0.49	..	***-0.129***	***0.053***	0.007
			0	0	..	..	***-0.056***	0.004
		0.8	0.18	0.16	0.031	***-0.758***	***0.140***	-0.036
			0.31	0.35	0.009	***-0.367***	***0.218***	-0.035
			0.52	0.49	0.009	***-0.185***	***0.219***	-0.020
			0.52	0	-0.008	..	***0.384***	-0.009
			0	0.49	..	***-0.212***	0.015	-0.008
			0	0	..	..	-0.016	-0.010
40	3	0.5	0.18	0.16	-0.032	***-0.606***	0.040	0.049
			0.31	0.35	0.016	***-0.115***	***0.055***	***-0.061***
			0.52	0.49	-0.028	***-0.094***	0.043	-0.022
			0.52	0	0.006	..	***0.094***	-0.042
			0	0.49	..	***-0.063***	0.001	0.017
			0	0	..	..	-0.009	0.002
		0.8	0.18	0.16	0.020	***-0.507***	***0.068***	0.006
			0.31	0.35	-0.024	***-0.184***	***0.122***	0.004
			0.52	0.49	0.020	***-0.173***	***0.122***	0.035
			0.52	0	-0.012	..	***0.210***	0.000
			0	0.49	..	***-0.103***	0.034	0.002
			0	0	..	..	-0.009	0.008
	5	0.5	0.18	0.16	***0.060***	***-0.142***	0.010	0.022
			0.31	0.35	0.027	***-0.101***	-0.002	-0.009
			0.52	0.49	0.010	***-0.063***	***0.113***	-0.018
			0.52	0	-0.004	..	***0.076***	0.000
			0	0.49	..	***-0.054***	0.028	-0.005
			0	0	..	..	-0.025	-0.007
		0.8	0.18	0.16	***-0.052***	***-0.500***	***0.130***	-0.004
			0.31	0.35	-0.024	***-0.148***	***0.107***	0.000
			0.52	0.49	0.024	***-0.129***	***0.124***	-0.008
			0.52	0	-0.021	..	***0.121***	-0.010
			0	0.49	..	***-0.116***	-0.003	-0.005
			0	0	..	..	-0.009	-0.007

Note. c = Number of clusters; m = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. .. indicates a true value of zero for the parameter. Bolded, italic values indicate relative parameter bias in excess of the recommended 0.05 cutoff (Hoogland & Boomsma, 1998).

For conditions with both mediated effects generated to be zero, the average *RPB* was 0.006, and -0.016 for conditions with a zero a_B and non-zero b_B and conditions with a zero a_B and zero b_B, respectively. The number of measurement occasions did not substantially impact recovery of the direct effect with average *RPB* values of 0.077 for 3 measurement occasions and 0.085 for 5 measurement occasions.

Relative Parameter Bias for Level-1 Variance Components

Table 10 presents the *RPB* for the level-1 variance components at each measurement occasion for the mediator. Remember that heteroscedasticity was assumed, and therefore, a level-1 residuals variance component was estimated for each measurement occasion for each outcome (mediator and distal outcome). Results indicate that the conditions that were problematic included 20 clusters and 3 measurement occasions. Within this set of conditions, the second measurement occasion showed no *RPB*. However, the first and last (third) measurement occasion showed substantial relative parameter bias for some conditions. The average *RPB* for the first, second, and third measurement occasions were 0.060, -0.005, and 0.047. In the conditions with unacceptable bias, all were over-estimated (positively-biased). The proportion of variance explained factor did not seem to influence the average *RPB*. For the first measurement occasions, the average *RPB* was 0.061 and 0.059 for the 0.5 and 0.8 proportions, respectively. For the third measurement occasions, the RPB values were 0.045 and 0.482 for the 0.5 and 0.8 proportions, respectively. No other conditions led to unacceptable *RPB* for the level-1 variance components.

Results were similar for the distal outcome's level-1 variance components. Only two of the conditions with 20 clusters and 3 measurement occasions led to unacceptable bias (see Table

Table 10

Relative Parameter Bias for the Mediator's Level-1 Residuals' Variance Component Estimates for Conditions with Three and Five Measurement Occasions

Design Conditions		Generating Parameter Values			Level-1 Variance for Mediator				
c	*m*	Prop EV	a_B	b_B	${}_M\sigma_0^2$	${}_M\sigma_1^2$	${}_M\sigma_2^2$	${}_M\sigma_3^2$	${}_M\sigma_4^2$
20	3	0.5	0.18	0.16	***0.060***	-0.004	0.036	..	..
			0.31	0.35	0.043	0.007	0.033	..	..
			0.52	0.49	***0.065***	-0.008	0.028	..	..
			0.52	0	***0.065***	-0.010	***0.058***	..	..
			0	0.49	***0.055***	0.009	***0.055***	..	..
			0	0	***0.075***	0.000	***0.059***	..	..
		0.8	0.18	0.16	***0.079***	0.004	0.025	..	..
			0.31	0.35	0.030	-0.001	***0.062***	..	..
			0.52	0.49	***0.057***	-0.020	0.034	..	..
			0.52	0	0.048	-0.017	***0.059***	..	..
			0	0.49	***0.076***	-0.007	***0.067***	..	..
			0	0	***0.064***	-0.012	0.042	..	..
	5	0.5	0.18	0.16	0.022	-0.008	0.003	0.000	-0.002
			0.31	0.35	0.011	0.005	-0.001	0.006	-0.003
			0.52	0.49	0.008	0.001	-0.002	0.012	0.021
			0.52	0	0.020	-0.002	0.000	-0.002	-0.008
			0	0.49	0.019	0.012	-0.003	0.018	0.001
			0	0	0.017	0.013	0.005	0.010	0.010
		0.8	0.18	0.16	-0.001	0.000	-0.005	0.000	0.009
			0.31	0.35	-0.001	-0.006	0.003	0.001	0.007
			0.52	0.49	0.002	0.011	0.002	0.000	0.003
			0.52	0	-0.006	0.009	0.010	0.006	0.000
			0	0.49	0.011	0.005	-0.007	0.008	-0.006
			0	0	0.009	0.006	0.003	0.006	0.000
40	3	0.5	0.18	0.16	0.035	-0.005	0.025	..	..
			0.31	0.35	0.019	-0.001	0.011	..	..
			0.52	0.49	0.018	0.000	0.007	..	..
			0.52	0	0.014	0.004	0.010	..	..
			0	0.49	0.030	0.007	0.010	..	..
			0	0	0.026	0.007	0.000	..	..
		0.8	0.18	0.16	0.018	-0.003	0.026	..	..
			0.31	0.35	0.034	-0.009	0.030	..	..
			0.52	0.49	0.044	-0.021	0.037	..	..
			0.52	0	0.006	-0.003	0.008	..	..
			0	0.49	0.040	-0.007	0.046	..	..
			0	0	0.030	-0.011	0.021	..	..
	5	0.5	0.18	0.16	0.015	0.002	0.002	0.006	0.005
			0.31	0.35	0.012	-0.002	0.000	0.001	0.005
			0.52	0.49	0.007	0.003	0.001	0.002	0.002
			0.52	0	0.017	0.004	0.001	0.008	0.008
			0	0.49	0.009	-0.001	0.008	0.000	0.000
			0	0	0.014	0.009	0.009	0.015	0.001
		0.8	0.18	0.16	0.010	-0.007	0.004	0.001	0.003
			0.31	0.35	0.018	0.003	0.007	0.007	0.005
			0.52	0.49	0.002	-0.010	0.005	0.007	-0.005
			0.52	0	0.016	-0.005	-0.002	0.004	0.012
			0	0.49	-0.013	0.008	-0.001	0.004	0.000
			0	0	-0.001	0.004	0.010	-0.010	0.005

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. .. indicates measurement occasions not modeled. Bolded, italic values indicate relative parameter bias in excess of the recommended 0.05 cutoff (Hoogland & Boomsma, 1998).

Table 11

Relative Parameter Bias for the Distal Outcome's Level-1 Residuals' Variance Component Estimates for Conditions with Three and Five Measurement Occasions

Design Conditions		Generating Parameter Values			Level-1 Variance for Distal Outcome				
c	*m*	Prop EV	a_B	b_B	${}_Y\sigma_0^2$	${}_Y\sigma_1^2$	${}_Y\sigma_2^2$	${}_Y\sigma_3^2$	${}_Y\sigma_4^2$
20	3	0.5	0.18	0.16	0.042	0.002	0.031	..	..
			0.31	0.35	***0.060***	0.005	0.020	..	..
			0.52	0.49	0.033	0.001	0.009	..	..
			0.52	0	0.036	0.015	0.022	..	..
			0	0.49	0.049	0.009	0.018	..	..
			0	0	0.038	0.015	0.015	..	..
		0.8	0.18	0.16	0.031	-0.003	0.021	..	..
			0.31	0.35	0.011	0.000	0.029	..	..
			0.52	0.49	-0.010	0.008	0.006	..	..
			0.52	0	0.028	0.006	-0.003	..	..
			0	0.49	0.022	-0.009	0.024	..	..
			0	0	***0.051***	-0.018	0.035	..	..
	5	0.5	0.18	0.16	0.006	0.003	-0.005	0.002	-0.004
			0.31	0.35	0.023	0.002	0.008	0.007	-0.003
			0.52	0.49	0.020	0.006	0.006	0.002	0.007
			0.52	0	0.025	0.007	0.007	0.006	0.002
			0	0.49	0.009	0.007	0.012	-0.005	-0.006
			0	0	0.026	0.009	0.009	-0.003	0.005
		0.8	0.18	0.16	0.000	0.002	0.003	0.004	0.003
			0.31	0.35	0.014	-0.004	-0.011	0.009	0.006
			0.52	0.49	-0.014	0.011	0.010	0.003	0.005
			0.52	0	-0.009	0.000	0.002	0.000	0.007
			0	0.49	0.010	0.002	0.004	-0.002	0.004
			0	0	0.003	0.011	0.004	-0.002	0.007
40	3	0.5	0.18	0.16	0.029	-0.003	0.012	..	..
			0.31	0.35	0.014	0.003	0.011	..	..
			0.52	0.49	0.019	0.003	0.014	..	..
			0.52	0	0.024	0.008	0.024	..	..
			0	0.49	0.041	-0.007	0.024	..	..
			0	0	0.027	0.003	0.014	..	..
		0.8	0.18	0.16	0.016	0.010	0.019	..	..
			0.31	0.35	0.012	-0.010	-0.001	..	..
			0.52	0.49	-0.009	0.000	0.000	..	..
			0.52	0	0.006	-0.016	0.006	..	..
			0	0.49	0.005	0.009	0.009	..	..
			0	0	0.014	-0.014	0.005	..	..
	5	0.5	0.18	0.16	0.005	0.005	0.007	0.001	0.003
			0.31	0.35	0.010	0.005	0.008	-0.002	0.008
			0.52	0.49	0.007	-0.004	-0.003	0.005	0.002
			0.52	0	0.002	0.005	-0.001	0.002	0.005
			0	0.49	0.010	0.009	-0.007	0.004	0.002
			0	0	0.004	0.001	0.002	-0.001	0.008
		0.8	0.18	0.16	-0.002	0.001	0.002	0.013	-0.007
			0.31	0.35	0.001	0.003	-0.005	0.000	-0.001
			0.52	0.49	-0.008	-0.002	0.004	0.002	-0.002
			0.52	0	0.000	0.006	0.010	-0.001	0.005
			0	0.49	0.004	0.000	0.005	0.002	-0.002
			0	0	0.000	0.003	0.003	0.002	-0.006

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. .. indicates measurement occasions not modeled. Bolded, italic values indicate relative parameter bias in excess of the recommended 0.05 cutoff (Hoogland & Boomsma, 1998).

11), and only for level-1 distal outcome residuals' variance estimates for the first measurement occasion.

Relative Parameter Bias for Intercept and Slope Fixed Effect

Table 12 contains the *RPB* for the fixed effect estimates of the mediator's intercept and slope and the distal outcome's intercept and slope. Two conditions for the distal outcome's overall slope showed slightly unacceptable *RPB* (with values of 0.058 and -0.055). Both conditions with unacceptable bias included 20 clusters with five measurement occasions, and different generating values for the proportion of explained variance.

Highest Posterior Density (HPD) Interval Coverage Rates

As noted previously, *HPD* intervals are the narrowest interval of values from a parameter's posterior distribution that includes values with the highest probability density. Credible intervals, which are equivalent to equal-tail intervals, ensures that the probability below the intervals equals the probability above. For example, a 95% *HPD* interval contains the narrowest interval containing 95% of the parameter's posterior distribution values whereas a 95% credible interval is the middle 95% of posterior distribution values. As a result, the HPD intervals are usually shorter in width than credible intervals.

Table 13 displays the percentage decrease in the width of the 95% HPD intervals compared to the 95% credible interval for the mediated effects. As shown in all but one case, there is a decrease in width when using the 95% *HPD* interval, suggesting a non-normal posterior distribution, though it may be slight. The width difference is less distinct with scenarios with more clusters. Due to the slightly narrower widths, *HPD* intervals are reported for all parameters of interest. Note that *HPD* intervals were provided by chain, and since three chains were run, the median of the percentages for the three chains was computed and used to provide

Table 12

Relative Parameter Bias for the Fixed Effects Estimates

Design Conditions		Generating Parameter Values			Intercept and Slope Fixed Effects			
c	m	Prop EV	a_B	b_B	${}_M\gamma_{000}$	${}_M\gamma_{100}$	${}_Y\gamma_{000}$	${}_Y\gamma_{100}$
20	3	0.5	0.18	0.16	0.028	-0.037	0.005	-0.001
			0.31	0.35	0.017	0.009	-0.013	0.046
			0.52	0.49	0.033	0.011	-0.013	0.024
			0.52	0	-0.006	0.026	0.010	-0.016
			0	0.49	-0.017	0.025	0.013	0.014
			0	0	-0.001	0.011	-0.006	-0.023
		0.8	0.18	0.16	0.006	-0.021	-0.009	-0.005
			0.31	0.35	-0.018	0.025	-0.010	0.027
			0.52	0.49	-0.004	-0.009	0.012	-0.007
			0.52	0	-0.005	0.007	-0.022	0.030
			0	0.49	0.011	0.023	0.003	-0.001
			0	0	0.003	-0.017	-0.009	0.013
	5	0.5	0.18	0.16	-0.008	0.016	-0.014	0.006
			0.31	0.35	0.032	-0.010	0.018	-0.029
			0.52	0.49	-0.026	0.024	-0.019	0.003
			0.52	0	0.010	0.023	-0.019	-0.010
			0	0.49	-0.008	0.004	0.019	-0.026
			0	0	-0.019	0.015	-0.011	***0.058***
		0.8	0.18	0.16	-0.007	0.012	0.003	-0.001
			0.31	0.35	0.014	-0.018	0.016	***-0.055***
			0.52	0.49	-0.005	0.013	0.001	0.012
			0.52	0	0.003	0.010	0.015	-0.002
			0	0.49	0.010	0.005	0.002	-0.015
			0	0	-0.001	-0.006	-0.013	-0.002
40	3	0.5	0.18	0.16	-0.012	0.007	-0.010	0.002
			0.31	0.35	-0.009	0.006	-0.011	-0.010
			0.52	0.49	0.000	-0.007	-0.008	0.001
			0.52	0	-0.020	-0.012	-0.003	-0.009
			0	0.49	0.002	-0.006	0.010	-0.039
			0	0	-0.001	-0.020	0.018	-0.012
		0.8	0.18	0.16	-0.010	0.006	-0.011	0.017
			0.31	0.35	-0.009	0.025	-0.012	0.016
			0.52	0.49	0.002	-0.002	0.002	0.000
			0.52	0	-0.010	-0.011	0.009	-0.025
			0	0.49	0.008	0.015	0.001	-0.027
			0	0	-0.009	0.015	0.013	0.012
	5	0.5	0.18	0.16	0.014	-0.034	-0.008	0.006
			0.31	0.35	0.008	-0.008	0.003	-0.003
			0.52	0.49	0.016	0.010	0.002	-0.018
			0.52	0	-0.006	0.006	0.003	0.000
			0	0.49	-0.012	0.004	0.002	-0.003
			0	0	0.017	0.008	-0.005	-0.001
		0.8	0.18	0.16	0.000	0.018	-0.002	0.003
			0.31	0.35	-0.001	0.001	-0.003	-0.015
			0.52	0.49	-0.009	0.000	0.002	0.006
			0.52	0	-0.002	0.003	-0.003	-0.009
			0	0.49	0.007	-0.016	0.001	-0.017
			0	0	0.012	-0.001	0.004	-0.017

Note. c = Number of clusters; m = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. Bolded, italic values indicate relative parameter bias in excess of the recommended 0.05 cutoff (Hoogland & Boomsma, 1998).

Table 13

Percentage Decrease in Coverage Interval Width Using the 95% HPD Intervals Compared to the 95% Credible Intervals

Design Conditions		Generating Parameter Values			Mediated Effect	
c	*m*	Prop EV	a_B	b_B	Cross-Level	Cluster-Level
20	3	0.5	0.18	0.16	8.57%	4.37%
			0.31	0.35	9.37%	2.99%
			0.52	0.49	9.88%	6.32%
			0.52	0	10.28%	5.15%
			0	0.49	7.53%	2.27%
			0	0	9.31%	6.41%
		0.8	0.18	0.16	3.32%	5.55%
			0.31	0.35	5.61%	4.23%
			0.52	0.49	6.30%	2.44%
			0.52	0	6.26%	2.25%
			0	0.49	2.55%	2.65%
			0	0	3.38%	1.42%
	5	0.5	0.18	0.16	1.19%	0.68%
			0.31	0.35	2.11%	2.55%
			0.52	0.49	1.86%	1.46%
			0.52	0	3.84%	1.34%
			0	0.49	2.63%	2.50%
			0	0	3.44%	1.13%
		0.8	0.18	0.16	3.57%	3.03%
			0.31	0.35	1.88%	3.68%
			0.52	0.49	2.28%	2.66%
			0.52	0	2.50%	2.86%
			0	0.49	1.66%	3.60%
			0	0	2.67%	2.47%
40	3	0.5	0.18	0.16	6.30%	5.48%
			0.31	0.35	10.09%	3.56%
			0.52	0.49	7.25%	3.26%
			0.52	0	8.95%	3.06%
			0	0.49	3.73%	2.39%
			0	0	2.00%	2.78%
		0.8	0.18	0.16	2.68%	2.78%
			0.31	0.35	3.16%	2.04%
			0.52	0.49	3.05%	2.04%
			0.52	0	2.28%	1.22%
			0	0.49	2.32%	2.03%
			0	0	1.48%	-0.32%
	5	0.5	0.18	0.16	3.54%	0.80%
			0.31	0.35	2.18%	3.63%
			0.52	0.49	1.92%	2.91%
			0.52	0	1.94%	5.80%
			0	0.49	1.70%	1.52%
			0	0	0.92%	2.43%
		0.8	0.18	0.16	1.22%	3.58%
			0.31	0.35	3.67%	1.08%
			0.52	0.49	0.51%	1.44%
			0.52	0	2.57%	0.83%
			0	0.49	1.46%	0.43%
			0	0	2.89%	1.15%

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance.

coverage rates. An acceptable criterion for each 95% *HPD* interval was defined using a lower limit of 92.48% up to an upper limit of 97.52% as described earlier with Equation 41.

Highest Posterior Density (HPD) Interval Coverage Rates for Mediated Effect Estimates

Table 14 provides the coverage rates for the mediated effects. The cluster-level mediated effect had coverage rates of either 100% or 99.67% for conditions with both indirect effects, a_B and b_B , generated to be zero. Conditions that had a generating value of zero value for a_B and a non-zero value for b_B, (hence a cluster-level mediated effect of zero), also resulted in overly high coverage rates. Conditions with the small, non-zero mediated effect had high, unacceptable converge rates for conditions with 20 clusters. As the number of clusters increased to 40, some of the conditions with a small, non-zero mediated effect showed acceptable coverage rates. Conditions with medium and large mediated effects had similar coverage rates, although the discrepancy between these two sets of conditions became smaller in scenarios with more clusters. In some conditions with 40 clusters, the two sets of conditions with medium and large mediated effects at the cluster-level had equal *HPD* interval coverage rates. All conditions that had a generating non-zero value for a_B and a zero b_B, and hence a cluster-level mediated effect of zero, showed acceptable coverage rates.

The cross-level mediated effect coverage rates are also reported in Table 14. Most of the conditions showed acceptable coverage rates. Three conditions with a zero cross-level mediated effect had overly high coverage rates. For all three of these conditions, the cluster-level mediated effect was also generated to be zero and had overly high coverage rates for the same conditions. Two conditions reported too-low coverage rates, with both conditions having the large value for a_B (0.52), and therefore, a large cross-level mediated effect.

Table 14

Percentage Coverage Rates of the 95% HPD Interval of the Mediated Effects' Estimates

Design Conditions		Generating Parameter Values			Mediated Effect	
c	*m*	Prop EV	a_B	b_B	Cross-Level	Cluster-Level
20	3	0.5	0.18	0.16	96.00	***99.00***
			0.31	0.35	95.67	***98.67***
			0.52	0.49	95.00	***91.67***
			0.52	0	96.33	96.33
			0	0.49	***98.33***	***99.67***
			0	0	***98.33***	***100.00***
		0.8	0.18	0.16	96.33	***99.33***
			0.31	0.35	93.67	***89.33***
			0.52	0.49	93.00	93.67
			0.52	0	94.67	93.67
			0	0.49	96.67	***99.00***
			0	0	95.67	***100.00***
	5	0.5	0.18	0.16	96.00	***99.67***
			0.31	0.35	95.67	96.33
			0.52	0.49	95.67	92.67
			0.52	0	94.00	97.00
			0	0.49	95.67	***99.67***
			0	0	94.33	***100.00***
		0.8	0.18	0.16	95.67	***98.00***
			0.31	0.35	93.67	***89.00***
			0.52	0.49	95.67	93.00
			0.52	0	93.33	96.00
			0	0.49	93.33	***99.00***
			0	0	96.67	***100.00***
40	3	0.5	0.18	0.16	95.00	***99.33***
			0.31	0.35	94.67	95.67
			0.52	0.49	***91.67***	95.33
			0.52	0	94.00	94.33
			0	0.49	***99.00***	***99.33***
			0	0	96.67	***99.67***
		0.8	0.18	0.16	94.00	95.00
			0.31	0.35	93.33	93.00
			0.52	0.49	94.67	93.00
			0.52	0	96.33	93.33
			0	0.49	96.33	***98.33***
			0	0	95.33	***99.67***
	5	0.5	0.18	0.16	93.67	***99.00***
			0.31	0.35	96.00	94.67
			0.52	0.49	94.67	94.67
			0.52	0	95.33	97.00
			0	0.49	94.67	***98.33***
			0	0	96.00	***100.00***
		0.8	0.18	0.16	92.67	93.67
			0.31	0.35	93.00	93.67
			0.52	0.49	96.00	***92.33***
			0.52	0	***92.00***	96.67
			0	0.49	96.67	***98.33***
			0	0	94.33	***99.67***

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. Bolded, italic values indicate unacceptable coverage rates outside criterion interval of 92.48% to 97.52% given 300 replications per conditions (Burton, Altman, Royston, Holder, 2006).

Table 15

Percentage Coverage Rates of the 95% HPD Interval of the Direct and Indirect Effects' Estimates

Design Conditions		Generating Parameter Values			Effects			
c	m	Prop EV	a_B	b_B	a_B	b_B	c'_B	b_W
20	3	0.5	0.18	0.16	96.33	***91.67***	95.33	93.67
			0.31	0.35	95.00	94.00	94.33	94.33
			0.52	0.49	94.67	92.67	93.00	94.33
			0.52	0	96.33	***92.00***	95.33	93.33
			0	0.49	93.00	93.67	96.00	93.00
			0	0	95.00	95.00	95.67	95.67
		0.8	0.18	0.16	95.00	94.00	95.33	96.33
			0.31	0.35	***92.33***	***91.33***	***92.33***	96.67
			0.52	0.49	95.33	92.67	95.00	97.00
			0.52	0	95.00	92.67	***91.33***	95.00
			0	0.49	95.00	92.67	95.67	96.33
			0	0	93.67	***91.67***	95.33	93.67
	5	0.5	0.18	0.16	95.00	93.33	94.00	95.67
			0.31	0.35	95.00	94.67	96.00	93.00
			0.52	0.49	96.67	92.67	92.67	93.67
			0.52	0	94.33	94.00	93.33	95.33
			0	0.49	94.33	96.33	95.00	94.67
			0	0	***92.33***	97.00	93.67	96.00
		0.8	0.18	0.16	95.67	93.33	94.67	94.00
			0.31	0.35	94.33	***91.00***	***88.33***	94.33
			0.52	0.49	95.33	93.33	93.33	96.33
			0.52	0	94.33	93.67	93.00	93.00
			0	0.49	93.33	***91.33***	93.00	***92.00***
			0	0	96.33	93.33	96.33	92.67
40	3	0.5	0.18	0.16	93.67	96.67	***92.00***	96.00
			0.31	0.35	93.67	96.00	95.33	92.67
			0.52	0.49	95.00	94.00	94.33	92.67
			0.52	0	94.67	***92.33***	94.33	93.33
			0	0.49	***92.00***	96.33	93.00	95.33
			0	0	94.00	93.00	96.67	96.00
		0.8	0.18	0.16	94.00	***90.67***	92.67	93.67
			0.31	0.35	94.00	94.00	94.00	93.33
			0.52	0.49	93.33	94.00	***92.33***	93.33
			0.52	0	95.67	93.67	***90.33***	95.33
			0	0.49	95.33	92.67	96.67	95.00
			0	0	94.33	***91.33***	95.67	94.67
	5	0.5	0.18	0.16	93.33	94.67	94.67	94.00
			0.31	0.35	95.33	96.00	96.00	95.33
			0.52	0.49	93.67	94.00	95.33	94.67
			0.52	0	97.33	96.67	93.33	93.33
			0	0.49	95.00	94.33	94.67	95.67
			0	0	94.67	96.67	***92.33***	94.67
		0.8	0.18	0.16	95.00	94.00	***91.33***	94.67
			0.31	0.35	93.33	94.67	***90.67***	96.00
			0.52	0.49	95.00	93.33	***92.00***	96.33
			0.52	0	93.00	95.33	93.33	93.33
			0	0.49	96.33	94.67	95.67	94.00
			0	0	93.33	93.67	95.33	94.67

Note. c = Number of clusters; m = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. Bolded, italic values indicate unacceptable coverage rates outside criterion interval of 92.48% to 97.52% given 300 replications per conditions (Burton, Altman, Royston, Holder, 2006).

Highest Posterior Density (HPD) Interval Coverage Rates for Direct and Indirect Effects' Estimates

The coverage rates for the direct effect (c'_B) and components of the indirect effects (a_B, b_B, and b_W) are reported in Table 15. The average *HPD* interval coverage rates were 94.87%, 94.06%, 94.51%, and 94.92% for a_B , b_B , c'_B , and b_W , respectively. Interval estimates of the b_B and c'_B parameters resulted in the most frequent occurrences of unacceptable coverage rates, which is consistent with the *RPB* results where these parameters also had the most conditions with unacceptable bias.

Highest Posterior Density (HPD) Interval Coverage Rates for Level-1 Variance Components

Table 16 and 17 display the *HPD* interval coverage rates for the level-1 residuals' variance components for the mediator and distal outcome, respectively. Too low coverage rates were most typically found in the conditions in which the coverage rates were unacceptable. Unacceptable coverage rates were found more often for the mediator's than for the distal outcome's level-1 residuals' variance component estimates In addition, most of these conditions occurred when there were 20 clusters and 3 measurement occasions; most also were found for the first (baseline) measurement occasion's level-1 variance. There were much fewer unacceptable coverage rates in the fourth and fifth measurement occasions for conditions with 5 measurement occasions.

Highest Posterior Density (HPD) Interval Coverage Rates for Fixed Effects' Estimates

Table 18 provides the HPD interval coverage rates for the fixed effects for the mediator and distal outcome. As can be seen in the Table, estimates of the mediator's slope resulted in the most conditions with unacceptable coverage rates.

Table 16

Percentage Coverage Rates of the 95% HPD Interval for the Mediator's Level-1 Residuals' Variance Components' Estimates

Design Conditions		Generating Parameter Values			Level-1 Variance for Mediator				
c	*m*	Prop EV	a_B	b_B	${}_M\sigma_0^2$	${}_M\sigma_1^2$	${}_M\sigma_2^2$	${}_M\sigma_3^2$	${}_M\sigma_4^2$
20	3	0.5	0.18	0.16	***89.00***	94.33	94.33	..	..
			0.31	0.35	92.67	96.00	94.33	..	..
			0.52	0.49	***89.00***	***92.33***	94.00	..	..
			0.52	0	94.33	96.67	93.67	..	..
			0	0.49	***92.33***	***91.00***	***89.33***	..	..
			0	0	***92.33***	94.00	92.67	..	..
		0.8	0.18	0.16	94.33	93.67	94.00	..	..
			0.31	0.35	***91.33***	95.00	93.33	..	..
			0.52	0.49	93.33	93.33	***92.00***	..	..
			0.52	0	***91.00***	94.00	***91.67***	..	..
			0	0.49	92.67	94.33	94.67	..	..
			0	0	93.00	95.67	95.33	..	..
	5	0.5	0.18	0.16	***92.00***	94.67	95.00	95.67	96.33
			0.31	0.35	93.33	97.00	95.33	96.67	95.00
			0.52	0.49	95.33	94.67	94.00	93.00	96.00
			0.52	0	95.67	93.00	94.33	94.33	94.00
			0	0.49	93.00	96.00	***91.67***	92.67	93.67
			0	0	93.00	95.00	94.67	94.67	94.00
		0.8	0.18	0.16	95.00	95.00	95.00	***92.33***	96.00
			0.31	0.35	94.00	93.33	95.67	94.00	96.33
			0.52	0.49	95.33	96.33	95.67	94.67	94.67
			0.52	0	96.33	93.67	95.00	93.67	96.33
			0	0.49	95.67	93.00	93.00	94.67	94.00
			0	0	92.67	94.67	95.00	95.33	93.00
40	3	0.5	0.18	0.16	94.00	96.00	94.67	..	..
			0.31	0.35	***92.00***	93.00	95.00	..	..
			0.52	0.49	93.33	96.33	94.33	..	..
			0.52	0	***92.33***	***92.00***	95.67	..	..
			0	0.49	92.67	95.33	96.33	..	..
			0	0	94.00	95.33	94.00	..	..
		0.8	0.18	0.16	95.33	94.33	***90.33***	..	..
			0.31	0.35	95.00	***92.00***	95.67	..	..
			0.52	0.49	***91.33***	92.67	94.33	..	..
			0.52	0	94.33	95.33	93.67	..	..
			0	0.49	93.00	96.00	95.33	..	..
			0	0	94.33	93.33	***92.33***	..	..
	5	0.5	0.18	0.16	95.33	95.67	94.67	93.67	95.33
			0.31	0.35	93.67	94.00	95.67	94.67	94.33
			0.52	0.49	95.67	94.33	94.67	93.33	95.67
			0.52	0	96.33	96.00	***92.00***	95.33	95.00
			0	0.49	95.00	95.00	95.67	92.67	94.67
			0	0	95.00	95.67	94.67	92.67	95.67
		0.8	0.18	0.16	93.33	95.00	95.67	94.33	94.00
			0.31	0.35	95.00	95.00	95.67	95.67	95.33
			0.52	0.49	96.67	96.33	94.00	92.67	95.67
			0.52	0	93.67	92.67	***92.33***	95.33	93.00
			0	0.49	95.00	96.00	94.67	95.00	94.67
			0	0	-0.001	0.004	0.010	-0.010	0.005

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. .. indicates measurement occasions not modeled. Bolded, italic values indicate unacceptable coverage rates outside criterion interval of 92.48% to 97.52% given 300 replications per conditions (Burton, Altman, Royston, Holder, 2006).

Table 17

Percentage Coverage Rates of the 95% HPD Interval for the Distal Outcomes' Level-1 Residuals' Variance Components' Estimates

Design Conditions		Generating Parameter Values			Level-1 Variance for Distal Outcome				
c	*m*	Prop EV	a_B	b_B	$_Y\sigma_0^2$	$_Y\sigma_1^2$	$_Y\sigma_2^2$	$_Y\sigma_3^2$	$_Y\sigma_4^2$
20	3	0.5	0.18	0.16	94.67	95.67	95.67	..	..
			0.31	0.35	94.00	97.00	94.00	..	..
			0.52	0.49	94.33	95.33	97.00	..	..
			0.52	0	93.33	95.33	96.67	..	..
			0	0.49	94.00	***91.33***	96.00	..	..
			0	0	94.67	94.33	95.00	..	..
		0.8	0.18	0.16	93.33	93.67	***92.33***	..	..
			0.31	0.35	94.67	97.33	95.00	..	..
			0.52	0.49	93.67	94.00	95.67	..	..
			0.52	0	95.67	93.67	92.67	..	..
			0	0.49	***90.00***	***91.33***	***92.33***	..	..
			0	0	93.33	93.67	93.67	..	..
	5	0.5	0.18	0.16	93.67	96.33	95.67	95.33	***92.33***
			0.31	0.35	94.00	94.33	93.67	94.00	95.67
			0.52	0.49	94.00	95.67	94.33	96.00	92.67
			0.52	0	96.33	96.33	95.33	97.33	95.33
			0	0.49	95.33	***92.00***	94.00	97.00	96.33
			0	0	94.00	93.67	95.00	96.67	95.67
		0.8	0.18	0.16	***92.00***	94.67	96.00	95.00	95.00
			0.31	0.35	94.00	94.33	95.67	95.33	96.33
			0.52	0.49	93.33	94.00	95.67	94.33	95.67
			0.52	0	96.00	93.00	95.33	93.33	95.67
			0	0.49	94.00	94.67	94.33	95.00	93.67
			0	0	95.67	94.67	93.33	93.00	95.67
40	3	0.5	0.18	0.16	***92.00***	95.00	96.00	..	..
			0.31	0.35	93.67	96.00	94.67	..	..
			0.52	0.49	93.67	93.67	95.67	..	..
			0.52	0	***92.33***	95.67	94.67	..	..
			0	0.49	95.33	97.00	95.00	..	..
			0	0	***92.00***	96.00	94.00	..	..
		0.8	0.18	0.16	95.00	94.00	94.67	..	..
			0.31	0.35	94.00	93.00	94.00	..	..
			0.52	0.49	92.67	***92.33***	95.67	..	..
			0.52	0	***90.00***	93.67	95.33	..	..
			0	0.49	93.33	93.33	***92.33***	..	..
			0	0	93.67	92.67	94.00	..	..
	5	0.5	0.18	0.16	93.00	93.00	95.33	93.67	94.33
			0.31	0.35	95.33	94.00	95.33	96.00	96.67
			0.52	0.49	94.00	92.67	***91.67***	94.33	93.00
			0.52	0	95.33	95.33	94.00	96.33	94.67
			0	0.49	93.33	96.00	94.33	***91.67***	94.00
			0	0	96.67	95.00	93.33	97.00	94.33
		0.8	0.18	0.16	96.00	95.33	95.67	***92.33***	93.67
			0.31	0.35	95.00	96.67	94.00	96.00	96.67
			0.52	0.49	***92.33***	92.67	96.33	95.67	96.67
			0.52	0	94.67	95.67	94.00	94.67	94.00
			0	0.49	95.33	95.67	97.33	95.33	93.67
			0	0	93.00	94.00	93.00	***92.00***	92.67

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. .. indicates measurement occasions not modeled. Bolded, italic values indicate unacceptable coverage rates outside criterion interval of 92.48% to 97.52% given 300 replications per conditions (Burton, Altman, Royston, Holder, 2006).

Table 18

Percentage Coverage Rates of the 95% HPD Interval of the Fixed Effects' Estimates

Design Conditions		Generating Parameter Values			Intercept and Slope Fixed Effects			
c	m	Prop EV	a_B	b_B	$_M\gamma_{000}$	$_M\gamma_{100}$	$_Y\gamma_{000}$	$_Y\gamma_{100}$
20	3	0.5	0.18	0.16	94.67	95.67	94.00	***98.67***
			0.31	0.35	96.33	94.00	95.33	96.33
			0.52	0.49	96.33	94.67	94.67	96.33
			0.52	0	95.33	95.33	95.00	94.00
			0	0.49	96.00	94.67	96.67	96.33
			0	0	96.67	93.33	95.67	95.67
		0.8	0.18	0.16	94.33	***91.33***	95.67	96.33
			0.31	0.35	96.00	94.67	96.67	94.33
			0.52	0.49	97.00	95.00	95.33	96.33
			0.52	0	95.67	97.00	96.00	95.67
			0	0.49	94.67	92.67	94.67	93.67
			0	0	96.33	93.67	95.33	96.33
	5	0.5	0.18	0.16	95.67	95.67	97.33	96.33
			0.31	0.35	96.33	94.00	***92.33***	***92.00***
			0.52	0.49	94.67	***97.67***	95.67	96.33
			0.52	0	95.33	***97.67***	97.00	96.33
			0	0.49	96.00	96.33	93.00	95.33
			0	0	94.00	94.00	95.00	94.33
		0.8	0.18	0.16	95.00	94.33	95.00	94.33
			0.31	0.35	92.67	96.33	96.67	95.67
			0.52	0.49	93.00	***91.67***	96.00	94.33
			0.52	0	93.00	96.00	96.33	93.67
			0	0.49	94.67	95.33	95.33	95.67
			0	0	94.67	94.00	95.33	97.33
40	3	0.5	0.18	0.16	97.33	94.67	93.67	96.67
			0.31	0.35	96.00	95.00	95.67	94.67
			0.52	0.49	95.00	95.33	96.00	96.33
			0.52	0	93.67	93.67	94.67	95.67
			0	0.49	94.33	93.33	97.00	94.33
			0	0	94.00	93.67	94.67	95.67
		0.8	0.18	0.16	95.00	95.00	94.33	96.00
			0.31	0.35	97.33	94.67	92.67	95.00
			0.52	0.49	94.33	***92.00***	94.33	93.00
			0.52	0	94.00	95.67	96.00	95.67
			0	0.49	94.33	95.33	95.33	95.00
			0	0	96.67	93.33	94.33	97.00
	5	0.5	0.18	0.16	95.33	***91.67***	95.00	97.00
			0.31	0.35	96.33	94.67	94.33	93.67
			0.52	0.49	94.67	95.00	95.33	94.33
			0.52	0	***98.00***	97.00	94.33	96.33
			0	0.49	94.67	94.00	93.67	93.67
			0	0	***97.67***	***92.33***	94.33	93.33
		0.8	0.18	0.16	95.00	95.00	93.67	95.00
			0.31	0.35	94.33	95.67	96.00	94.00
			0.52	0.49	95.67	93.00	93.67	95.00
			0.52	0	97.00	94.67	92.67	96.00
			0	0.49	93.33	94.00	95.67	94.00
			0	0	95.67	93.33	92.67	97.00

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance. Bolded, italic values indicate unacceptable coverage rates outside criterion interval of 92.48% to 97.52% given 300 replications per conditions (Burton, Altman, Royston, Holder, 2006).

Statistical Power for Mediated Effects

Figure 5 provides the empirical power curves for the cross-level and cluster-level mediated effects. Conditions with a small cross-level mediated effect and 20 clusters were found to have low power regardless of the number of measurement occasions or true proportion of explained variance. However as the mediated effect increased, power increased substantially for all conditions except the condition with 3 measurement occasions and a smaller proportion of explained variance (0.5). The other three conditions showed similar increases when going from a small to medium mediated effect. The conditions with the larger proportion of variance (0.8), regardless of the number of measurement occasions, had power close to 0.80 for the medium mediated effect. When the cross-level mediated effect was large, power was greater than 0.80 in the condition with 5 measurement occasions and a smaller proportion of explained variance and nearly or equal to 1.0 for the conditions with the larger proportion of explained variance, regardless of the number of measurement occasions. Having a zero or non-zero cluster-level mediated effect simultaneously being estimated did not impact power. Power for the condition with 3 measurement occasions and smaller proportion of explained variance did not vary much across the true values of mediated effect sizes.

For 40 clusters, power was generally greater than the corresponding conditions with 20 clusters. While none of the conditions had power over 0.8 for the small mediated effect, three of the conditions had power over 0.8 for the medium effect, which entailed conditions with either a larger proportion of explained variance or a smaller proportion of explained variance paired with the larger number of measurement occasions. These same conditions had power close to or equal to 1.0 for the large mediated effect. As was shown with 20 clusters, the value of the cluster-level effect, whether zero or non-zero, did not influence power for the large cross-level mediated effect. Power for the condition with 3 measurement occasions, smaller proportion of explained variance, and 40 clusters had a larger increase in power going from the small to the medium mediated effect, compared to the same condition with 20 clusters. For both sets of conditions (20 clusters and 40 clusters), the gap between the conditions with 3 or 5 measurement occasions paired with a proportion of explained variance value of 0.5 was much greater than the gap between the conditions with 3 or 5 measurement occasions paired with a proportion of explained variance of 0.8.

For the cluster-level mediated effect's power analysis, results shown in Figure 5 show that there is little difference in power across any of the mediated effect sizes (i.e. the graph lines are closer together). With 20 clusters, the lines are the closest (compared to any of the other graphs) and the change between mediated effect sizes is also very similar across the conditions. No condition has power greater than 0.30 across the mediated effect sizes. However for the conditions with 40 clusters, the change between mediated effects becomes more prominent, even though power is still less than 0.70 for the large mediated effect. In addition, the same conditions that showed substantial increase across mediated effects' sizes for the cross-level mediated effect (all conditions except the condition with 3 measurement occasions and a proportion of 0.5) start to separate more noticeably from the fourth condition.

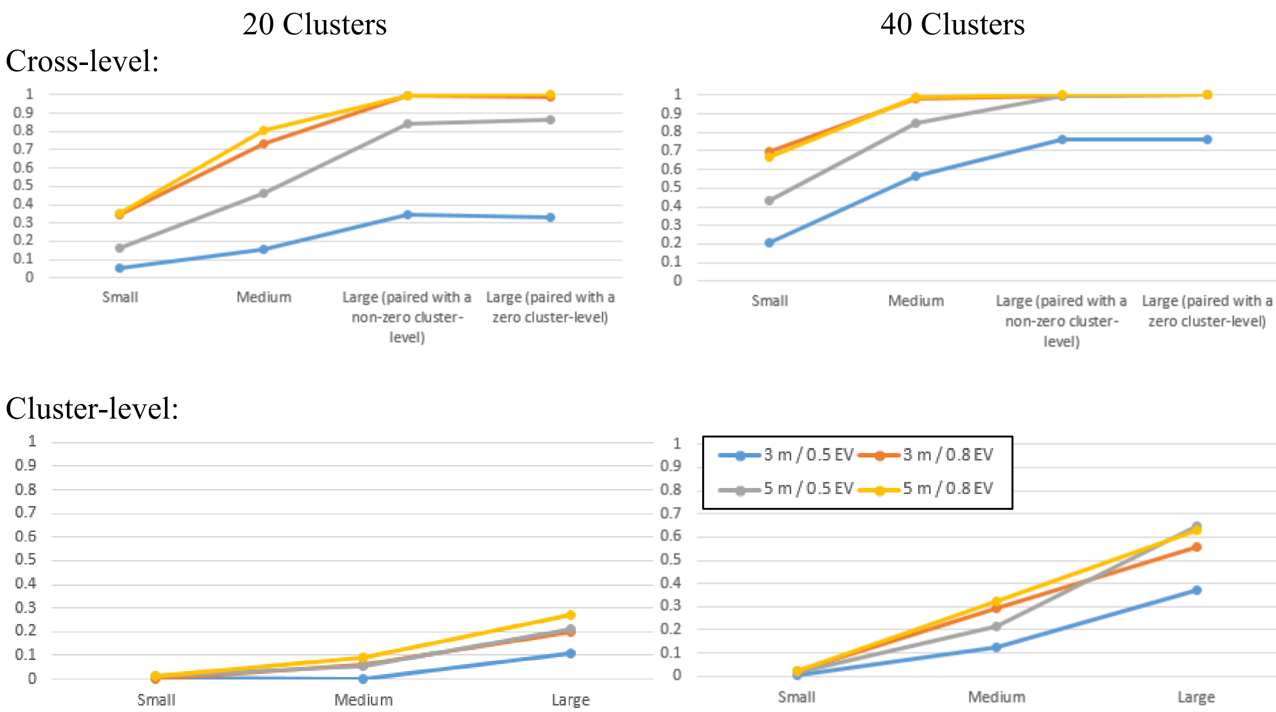

Figure 5. Empirical power for the cross-level and cluster-level mediated effect presented by the number of clusters, number of measurement occasions, proportion of explained variance, and mediated effect size. Top row contains the empirical power for the cross-level mediated effect, and the bottom row contains the power for estimation of the cluster-level mediated effect. The first column represent conditions with 20 clusters, and the second row represents conditions with 40 clusters. The horizontal axis distinguishes the true value for the mediated effect size.

Real Data Analysis

The model in Equations 30 through 32 was estimated using a subset of ECLS-K data. Variable *X* represented whether the student's kindergarten school had a full- versus part-time kindergarten. Reading IRT scores (Reardon, 2007) in fall and spring of kindergarten and spring, 1st grade provided the three mediator variable measures. Math IRT scores in spring 1st, 3rd and 5th grades constituted the distal outcome score. *Time* for *M* was coded with 0, 1 and 3 and with 0, 4, 8 for *Y*. Means and medians of the posterior distribution were captured. Note, however, the means and medians of the posterior distributions did not differ substantially for each parameter. Thus, the mean was used as each parameter's point estimate. The 95% highest posterior density

(*HPD*) interval for each parameter of interest (including the mediated effects, $a_B b_W$ and $a_B b_B$ and direct effect, c'_B) were captured and reported.

Note that the real data analysis was conducted only for methodological demonstration purposes. Broader generalization of these results for kindergarten and links to reading and math trajectories should not be made based on this analysis.

The mediator's intercept and slope were estimated to be -1.22 and 0.46, respectively, and 0.15 and 0.12 for the distal outcome supporting positive growth for reading and math scores (see Table 19). The direct effect (c'_B) estimate was −0.005 supporting a slightly negative effect of full- versus part-time kindergarten on mathematics growth after controlling for its mediated effect via reading growth. The within-school (student) "effect", b_W , of reading growth on mathematics growth was estimated to be -0.0004 representing the decrease predicted for a student's mathematics score growth given a one-point difference in the student's reading score growth, though this parameter was not statistically significant. However, for these data, the cluster-level effect, b_B , of the school's reading slope on the school mathematics slope was estimated to be -0.04 and was statistically significant ($p < .05$). The effect of full- versus part-time kindergarten on growth in reading was not statistically significant (a_B = -0.004). The resulting cluster-level and cross-level mediated effects ($a_B b_B$ and $a_B b_W$, respectively) had 95% *HPD* intervals which both included zero [with *HPD* intervals of (-0.0001,0.001) and (-0.00007, 0.00007), respectively].

Table 19
Fixed Effects Parameter, Standard Error and 95% HPD Interval Estimates

Parameter		Coeff.	Est.	*SE*	95% *HPD*
Model for Reading (Mediator) Initial Status					
	Intercept	${}_{M}\gamma_{000}$	-1.227	0.011	(-1.247, -1.204) *
Model for Reading (Mediator) Slope					
	Intercept	${}_{M}\gamma_{100}$	0.465	0.003	(0.459, 0.470) *
	X	a_B	-0.004	0.003	(-0.010, 0.002)
Model for Math (Outcome) Initial Status					
	Intercept	${}_{Y}\gamma_{000}$	0.157	0.008	(0.142, 0.174) *
Model for Math (Outcome) Slope					
	Intercept	${}_{Y}\gamma_{100}$	0.127	0.001	(0.125, 0.129) *
	X	c'_B	-0.005	0.001	(-0.008, -0.003) *
	Reading Slope	b_B	-0.049	0.125	(-0.089, -0.010) *
Model for Within-School *b* Parameter					
	Intercept	b_W	-0.000	0.006	(-0.013, 0.012)
Mediated Effects					
	Cluster-level	$a_B b_B$	0.0002	0.0001	(-0.0001, 0.001)
	Cross-level	$a_B b_W$	0.0000	0.0000	(-0.00007, 0.00007)

Note. Coeff. = Coefficient; Est. = Parameter estimate; *SE* = Standard error estimate; *HPD* = Highest posterior density interval. Means and medians of the posterior distributions did not differ substantially; means are reported.
**HPD* interval does not contain a value of zero.

The level-1 variance component estimates were larger at the first two measurement occasion for the reading versus mathematics achievement scores (see Table 20). As students continued through elementary school, the within-student variances of the scores at each time point decreased for both the mediator and the distal outcome, with the exception of a slight increase in the variance between the 1st and 2nd measurement occasion for the mediator. By the 3rd measurement occasion, the within-student variances had decreased substantially to 0.008 for both the mediator and distal outcome and were almost equivalent. All estimates were statistically significant.

Table 20

Level-1 Random Effects Variance Component Parameter and 95% HPD Interval Estimates

Measure and Time	Coefficient	Estimate	95% *HPD*
Reading (Mediator)			
Fall, Kinder	${}_M\sigma_0^2$	0.050	(0.046, 0.055) *
Spring, Kinder	${}_M\sigma_1^2$	0.052	(0.049, 0.055) *
Spring, 1st Grade	${}_M\sigma_3^2$	0.008	(0.003, 0.014) *
Math (Outcome)			
Spring, 1st Grade	${}_Y\sigma_0^2$	0.044	(0.042, 0.046) *
Spring, 3rd Grade	${}_Y\sigma_4^2$	0.035	(0.033, 0.036) *
Spring, 5th Grade	${}_Y\sigma_8^2$	0.008	(0.006, 0.011) *

Note. *HPD* = Highest posterior density interval; Means and medians of the posterior distributions did not differ substantially; means are reported.
* *HPD* interval does not contain a value of zero.

As discussed previously, the error variances may be a function of a predictor, such as *Time.* The results of this analysis indicates that the error variances appeared to decrease in value over time. To test whether there were statistically significant differences in level-one variances across measurement occasions, posterior distributions of the differences in these variances for each pair of measurement occasion within outcome type were captured. Table 21 contains point and interval estimates of these differences which support the existence of heteroscedasticity as a function of measurement occasion. As shown in the Table, all but one pair of mediating variable level-1 residuals' variances and all pairs of outcome variable level-1 residuals' variances differed significantly. Further, the computed proportion of explained variance to total variance generally increased over time, as a researcher would hope might occur when a treatment or invention takes effect. The proportions of explained level-1 variances for the mediator were 0.74, 0.70, and 0.93 and for the distal outcome were 0.62, 0.68, and 0.90 for the first, second, and third measurement occasion, respectively.

Table 21

Differences in Level-1 Random Effects Variance Component Parameters and 95% HPD Interval Estimates for the Differences

Measure and Time	Coefficients	Estimate	95% *HPD*
Reading (Mediator)			
Fall, Kinder vs Spring, Kinder	${}_{M}\sigma_0^2$ vs ${}_{M}\sigma_1^2$	0.006	(-0.008, 0.021)
Fall, Kinder vs Spring, 1st Grade	${}_{M}\sigma_0^2$ vs ${}_{M}\sigma_3^2$	-0.141	(-0.172, -0.111)*
Spring, Kinder vs Spring, 1st Grade	${}_{M}\sigma_1^2$ vs ${}_{M}\sigma_3^2$	-0.147	(-0.185, -0.112)
Math (Outcome)			
Spring, 1st Grade vs Spring, 3rd Grade	${}_{Y}\sigma_0^2$ vs ${}_{Y}\sigma_4^2$	-0.024	(-0.031, -0.017) *
Spring, 1st Grade vs Spring, 5th Grade	${}_{Y}\sigma_0^2$ vs ${}_{Y}\sigma_8^2$	-0.114	(-0.130, -0.098) *
Spring, 3rd Grade vs Spring, 5th Grade	${}_{Y}\sigma_4^2$ vs ${}_{Y}\sigma_8^2$	-0.090	(-0.107, -0.076) *

Note. *HPD* = Highest posterior density interval; Means and medians of the posterior distributions did not differ substantially; means are reported.
* *HPD* interval does not contain a value of zero.

The student level (level-2) intercept and slope residuals' variances and covariances for reading and mathematics can be found in Table 22. Remember that the two measures were both IRT-scaled and so the associated scores (and growth trajectory parameters) are expected to have a small-scale. The size of the reading intercept variance is the strongest indicating a lot of variability in the fall of kindergarten reading scores. However except for the 95% *HPD* intervals representing the covariance between the student's reading growth with mathematics' intercept, none of the other 95% *HPD* intervals contain zero, and all other level-2 random effects variances and covariances differ significantly from zero. Table 23 (below the diagonal) contains correlations (corresponding with standardized covariances) between the mediator and outcome's intercept and slope. The strongest correlation was found between the reading's intercept for the fall of kindergarten score and mathematics intercept for the spring, 1st grade score (r=0.731). The next strongest correlation was found between the reading intercept and reading slope (r=-0.448).

Table 22

Level-2 Random Effects Variance and Covariance Parameter and 95% HPD Interval Estimates

Parameter	Coefficient	Estimate	95% *HPD*
Variance			
Reading Int.	${}_M\tau^2_{\pi 0}$	0.143	(0.136, 0.150) *
Reading Slope	${}_M\tau^2_{\pi 1}$	0.007	(0.006, 0.008) *
Math Int.	${}_Y\tau^2_{\pi 0}$	0.072	(0.068, 0.076) *
Math Slope	${}_Y\tau^2_{\pi 1}$	0.000	(0.000, 0.000)*
Covariance			
Reading Int. with Reading Slope	${}_M\tau_{\pi 01}$	-0.014	(-0.016, -0.013) *
Reading Int. with Math Int.	${}_{M,Y}\tau_{\pi 00}$	0.074	(0.070, 0.079) *
Reading Int. with Math Slope	${}_{M,Y}\tau_{\pi 01}$	-0.001	(-0.001, -0.001) *
Reading Slope with Math Int.	${}_{M,Y}\tau_{\pi 10}$	0.000	(-0.001, 0.001)
Reading Slope with Math Slope	${}_{M,Y}\tau_{\pi 11}$	..	..
Math Int. with Math Slope	${}_Y\tau_{\pi 01}$	0.001	(0.001, 0.002) *

Note. *HPD* = Highest posterior density interval; Int. = Intercept; Math = Mathematics; .. = not estimated (constrained to zero). Means and medians of the posterior distributions did not differ substantially; means are reported.

**HPD* interval does not contain a value of zero.

Table 23

Estimates of Correlations Between Level-2 and Level-3 Residuals in Real Data Analysis

	Reading		Math	
Outcome and Parameter	Int.	Slope	Int.	Slope
Reading Int.		-0.557	0.777	-0.059
Reading Slope	-0.448		-0.109	..
Math Int.	0.731	0.004		-0.161
Math Slope	-0.187	..	0.342	

Note. Int. = Intercept; Math = Mathematics; .. = not estimated (constrained to zero). Correlations between pairs of level-2 residuals appear below the diagonal; correlations between pairs of level-3 residuals appear above the diagonal.

Table 24 contains the level-3 random effects covariance matrix elements. As at the student level, the school level variance for the reading intercept had the largest variance value $({}_M\hat{\tau}^2_{\beta 0} = 0.058)$. The *HPD* interval estimate of the covariance between the school's reading intercept and mathematics slope included zero (-0.001, 0.0000), and the *HPD* interval estimate of the covariance between the school's reading slope and mathematics intercept included zero (-0.002, 0.000). The latter was also found not to be statistically significant at the student level. The associated correlation values appear above the correlation matrix's diagonal in Table 23. As with the student-level correlations, the strongest correlation at the school level was found between the intercept in reading and mathematics intercept (r=0.777), and the next strongest (negative) correlation was found between the intercept and growth in reading (r=-0.557). The correlation between the mathematics intercept and mathematics slope was positive at the student-level (r=0.34), yet negative at the school-level (r=-0.16).

Table 24

Level-3 Random Effects Variance and Covariance Parameter and 95% HPD Interval Estimates

Parameter		Coefficient	Estimate	95% *HPD*
Variance				
	Reading Int.	${}_{M}\tau^2_{\beta 0}$	0.058	(0.050, 0.067) *
	Reading Slope	${}_{M}\tau^2_{\beta 1}$	0.002	(0.002, 0.003) *
	Math Int.	${}_{Y}\tau^2_{\beta 0}$	0.031	(0.027, 0.036) *
	Math Slope	${}_{Y}\tau^2_{\beta 1}$	0.000	(0.000, 0.000)*
Covariance				
	Reading Int. with Reading Slope	${}_{M}\tau_{\beta 01}$	-0.006	(-0.008, -0.005) *
	Reading Int. with Math Int.	${}_{M,Y}\tau_{\beta 00}$	0.033	(0.027, 0.039) *
	Reading Int. with Math Slope	${}_{M,Y}\tau_{\beta 01}$	-0.000	(-0.001, 0.000)
	Reading Slope with Math Int.	${}_{M,Y}\tau_{\beta 10}$	-0.001	(-0.002, 0.000)
	Reading Slope with Math Slope	${}_{M,Y}\tau_{\beta 11}$	..	..
	Math Int. with Math Slope	${}_{Y}\tau_{\beta 01}$	-0.000	(-0.001, -0.000) *

Note. *HPD* = Highest posterior density interval; Int. = Intercept; Math = Mathematics; .. = not estimated (constrained to zero). Means and medians of the posterior distributions did not differ substantially; means are reported.

[*] *HPD* interval does not contain a value of zero.

Chapter 5: Discussion

The current study's intent was to investigate recovery of mediated effects in a three-level longitudinal mediation model in which it is hypothesized that the impact of a cluster-level treatment affects growth in a distal outcome that is mediated by growth in a mediator for data including clusters of individuals. The study builds upon previous research from Cheong (2011) where a two-level longitudinal mediation model was simulated usingthe SEM framework, and a mediated effect was evaluated at the individual level. This study extended that model to include three levels (to handle clustered individuals) and simulated two mediated effects, a cross-level and cluster-level effect, with a shared effect (path a_B) for the treatment's effect on the mediating variable. In addition, analyses were conducted using the HLM framework to allow for further research on extensions of this model to include more complicated data structures, such as cross-classified or multiple-membership data.

Four factors were manipulated in the simulation study presented in this study: number of clusters, true value for the mediated effect, number of measurement occasions, and proportion of explained level-one variance. These factors were selected based on prior research on mediation in single-level and multilevel data. For instance, MacKinnon et al. (2002) found in their single-level mediation research that sample size and effect size decreased the relative parameter bias of estimates of mediated effects as these factors increased. Pituch and Stapleton (2008) found that increasing sample size and true effect size values lead to better results with power and coverage rates in their multilevel mediation analysis. Cheong's (2011) findings showed that increasing the proportion of explained level-1 variance improved the relative parameter bias of the mediated effect at level-two. Furthermore, research shows that increasing the number of measurement occasions improves precision of the estimates of the growth rate and statistical power (Muthen &

Curran, 1997; Singer & Willet, 2003). Results from this study further support that these four factors are important in decreasing the relative parameter bias and improving statistical power for estimates of indirect effects in mediation models and, more specifically from this study, in longitudinal mediation models.

The current study paralleled Cheong's (2011) study in several of the design parameters that were used, and results were found to be similar. Cheong evaluated six level-two sample sizes ranging from 100 to 5,000. Due to high *RPB* in the low sample sizes (100 and 200) and low *RPB* in the larger sample sizes (2,000 and 5,000), the current study purposefully focused on the middle sample sizes (500 and 1,000), as these sample sizes are also very typical in multilevel educational field studies (see, for example, Ivers et al., 2011 and Spybrook, 2007). Cheong evaluated a two-level model where the mediated effect occurred at the individual level, and therefore, the cross-level mediated effect results from this study may be more appropriate for comparison, as one component of this effect occurred at the level of the individual. One difference to remember between these two studies is that the treatment was assigned at the participant's level in Cheong's study, while, in the current study, intervention was assigned at the cluster level to replicate cluster randomized trials. Nevertheless, similar results were found for estimates of the mediated effects. In both studies, the proportion of explained level-1 variance, number of clusters, size of the mediated effect, and number of measurement occasion influenced the relative parameter bias of estimates of the mediated effect and its statistical power. Similar patterns across results were also seen in both studies. For instance, the mediated effect's relative parameter bias difference between corresponding conditions with different measurement occasions was larger when the proportion of explained variance was smaller in Cheong's and the current study. The difference in relative parameter bias decreased as the number of measurement occasions increased. In Cheong's study, acceptable relative parameter bias became more prevalent in sample sizes of 1,000 and larger, as was also found in the current study with conditions including 40 clusters with 25 individuals in each cluster for a total of 1,000 individuals. Both studies showed that improving the proportion of explained variance provided mostly accurate mediated effect estimates for many conditions, and seemed to have more impact on reducing the biases and improving power than increasing the number of measurement occasions. The improvement in statistical power across factors was generally the same between the studies; however, in some cases, power was greater in the SEM model compared to the current study given similar design parameter values. This was shown mainly in conditions with 20 clusters (or 500 sample size, in Cheong's study), 3 measurement occasions, and a proportion of explained variance of 0.5 or conditions with a small mediated effect. The addition of a cluster level may have resulted in less power due to more uncertainty in the model as more parameters were estimated. Furthermore for the cross-level mediated effect, the effect of the treatment on the mediator occurred at the cluster level, while the effect of the mediator's growth on the distal outcome's growth occurred at the individual level, which may have also contributed to its lower power. The independent variable was modeled as dichotomous in the current study and continuous in Cheong's study; however results were still shown to be similar. This is consistent with findings from previous research comparing continuous and dichotomous independent variables, and results are shown to be close to equivalent (MacKinnon et al., 2002).

Some differences between the studies include the direction in which the mediated effects were mostly biased. The cross-level (and cluster-level) mediated effects in the current study were mostly negatively-biased, while the mediated effect in Cheong's (2011) study was mostly positively-biased. This may be due to the complexity of also modeling a cluster-level mediated effect, along with its variances and covariances. In addition, the direct effect parameter at level-three, c'_B, was mostly positively-biased, while effect b_B was mostly negatively-biased in the current study which may have led to under-estimation of the cluster-level effect. In addition, the cross-level and cluster-level mediated effect in this study shared path a_B, which may have also partly influenced the direction of the biases that occurred for both mediated effects. Importantly, Cheong's model was estimated using maximum likelihood, while the current model used Bayesian MCMC estimation. Using different estimation methods, such as Bayesian inference versus maximum likelihood or restricted maximum likelihood inference, has been shown to impact parameter estimation differently resulting in differences in coverage rates and relative parameter bias for some parameters (Beerli, 2005; Celeux, El Anbari, Marin, Robert, 2015; Liu, Yu, Kalavacharla, & Liu, 2011; Localio, Berlin, Have, 2005; Yuan and MacKinnon, 2009). For example, in Yuan and MacKinnon's (2009) research on single-level mediation, there were 9 out of 20 conditions in which the biases of the indirect effect were in opposite directions when comparing a frequentist analysis with a Bayesian analysis with normal priors, although no estimation method favored a certain bias (positive versus negative) direction in this study. Further research would need to be conducted to evaluate this discrepancy.

The conditions that had a true value of zero for the mediated effects resulted in either acceptable coverage rates or, in some cases for the cluster-level mediated effect, over-coverage. The Type I error rates for the cluster-level mediated effect were slightly lower for conditions when path a_B was equal to zero and paired with a non-zero path b_B compared to when path b_B was equal to zero paired with a non-zero path a_B. This was inconsistent with previous literature. MacKinnon et al. (2002) found for that some statistical mediation tests that proved to be optimal for non-zero mediated effects or mediated effects where both paths, a and b, were zero, were problematic and produced high Type I error rates for conditions when only one parameter for the indirect effect was equal to zero. In some previous studies, the Type I error rates were higher when the mediated effect had a zero value for path *a* paired with a high value of path *b* compared to vice-versa (MacKinnon et al., 2002; Pituch & Stapleton, 2008). The coverage rates in this study were computed based on using the highest posterior density intervals, which do not require parametric assumptions for the posterior distribution's shape. The ability to relax certain assumptions when constructing the interval may have resulted in better (and even shorter) intervals that included the true value. When both paths, a_B and b_B, were generated to be zero, the Type I error rates were extremely low. These findings were consistent with mediation research that found very low Type I error rates when both paths in an indirect effect were generated to be zero (MacKinnon et al., 2002; 2004; Pituch et al., 2006; Yuan and MacKinnon, 2009).

Increasing the effect size and number of clusters did lead to improvement in the relative parameter bias for the mediated effects. The expectation that the effect size of mediated effects and sample size are important factors in increasing the accuracy of mediated effects is consistent with findings in prior studies (see MacKinnon et. al., 2002). However, the empirical coverage rates mostly decreased as the effect size of the mediated effect increased. Only in conditions that had a large number of clusters and measurement occasions did coverage rates for the cross-level mediated effect begin to stabilize or even increase as effect size increased.

Differences between the cross-level and cluster-level mediated effects' outcome measures are consistent with Pituch and Stapleton's research (2012) in which the authors evaluated cross-level and cluster-level mediation processes in two-level cross-sectional mediation models. Both studies showed that increasing the number of clusters and effect size improved power for the mediated effects, and the coverage rate for the cluster-level mediated effect was overestimated when the effect was small. Furthermore, Pituch and Stapleton found that the cross-level mediated effect had much greater power than the cluster-level mediated effect in models that were simulated with a contextual effect. In some cases, power for the cross-level was as high as almost 28 times the cluster-level mediated effect for conditions with small effect sizes and number of clusters. The current study also found substantially higher power for the cross-level mediated effect compared to the cluster-level mediated effect.

As sample size was shown to substantially impact recovery of the mediated effects in the simulation study, an additional small-scale pilot simulation study was conducted using the four conditions presented in Table 25. In this added study, the number of clusters was substantially increased to 100 (to compare with 20 and 40). In addition, only 70 converged replications were used to perform the analysis. Restricting this small-scale add-on analysis to this small number of replications was due to the extremely long time required for estimation of the model for each dataset given the larger number of clusters in each dataset. It should be noted that these conditions include small effect sizes, and results described later for this pilot simulation study are expected to improve as the mediated effect size increases.

Table 25

Pilot Simulation Design Conditions and Generating Parameter Values

Design Conditions		Generating Parameter Values				
					Mediated Effect	
c	m	Prop EV	a_B	b_B	Cross-Level	Cluster-Level
100	3	0.5	0.18	0.16	0.0630	0.0288
		0.8	0.18	0.16	0.0630	0.0288
	5	0.5	0.18	0.16	0.0630	0.0288
		0.8	0.18	0.16	0.0630	0.0288

Note. Generating value for b_W was 0.35 across all conditions and 25 level-2 units per level-3 unit were generated. c = Number of clusters; m = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance.

Table 26 reports the percentage of converged replications for each condition. As was similarly shown in Table 6, for conditions with 20 and 40 clusters, the condition with a small number of measurement occasions and a small proportion of explained variance, 3 and 0.5 respectively, was found to be problematic, although improvement in convergence rates was evident for data that included more clusters. The convergence rate for 3 measurement occasions and a proportion of explained variance of 0.5 was 19.52% and 22.10% for 20 and 40 clusters, respectively (see Table 6) and the corresponding rate for 100 clusters was 69.33%. Clearly, at least 100 and likely more clusters are needed to improve likelihood of convergence when estimating this model.

Table 26

Convergence Rates Across Pilot Simulation Conditions with 100 Clusters

Manipulated Factors			Convergence Rates
c	*m*	Prop EV	Across First Set of 70 Replications[a]
100	3	0.5	69.33%
100	3	0.8	100.00%
100	5	0.5	100.00%
100	5	0.8	100.00%

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance.
[a]Subsequent replication datasets were generated to replace non-converged replications.

Table 27 reports the relative parameter bias for estimates of the cross-level and cluster-level mediated effects across conditions. Improvement in the relative parameter bias for the cluster-level mediated effect was shown with the increase in the number of clusters. From Table 7, the lowest relative parameter bias for the cluster-level effect with 40 clusters and a small, non-zero mediated effect was -0.570 compared to the highest relative parameter bias of -0.285 using 100 clusters. As was found with 20 and 40 clusters, the cluster-level is still strongly negatively-biased for some of the conditions. One condition had acceptable relative parameter bias, which did not occur in any of the conditions with 20 or 40 clusters. The only condition that showed unacceptable relative parameter bias for the cross-level mediated effect was found in the condition with the small number of measurement occasions and a small proportion of explained variance. The conditions with these design parameters (3 measurement occasions paired with 0.5 proportion of explained variance) were shown to also be problematic with 20 and 40 clusters when estimating the cross-level mediated effect. An applied researcher should keep this in mind when deciding on the number of measurement occasions to capture when estimating the cross-level mediated effect. Researchers should also explore ways to enhance the likelihood that there is a high proportion of explained level-1 variance such as by using highly reliable measures.

Table 27

Relative Parameter Bias for the Non-Zero Mediated Effects' Estimates

Design Conditions		Generating Parameter Values			Mediated Effect	
c	*m*	Prop EV	a_B	b_B	Cross-Level	Cluster-Level
100	3	0.5	0.18	0.16	***-0.186***	***-0.180***
		0.8	0.18	0.16	0.015	***-0.285***
	5	0.5	0.18	0.16	0.027	***-0.203***
		0.8	0.18	0.16	-0.020	-0.021

Note. *c* = Number of clusters; *m* = Number of measurement occasions per a student; Prop EV = Proportion of explained level-1 residuals' variance; .. indicates a true value of zero for the cluster-level mediated effect paired with a non-zero cross-level mediated effect. Bolded, italic values indicate relative parameter bias in excess of the recommended 0.05 cutoff (Hoogland & Boomsma, 1998).

While increasing the number of clusters to, for instance 100, substantially reduces the relative parameter bias for the cluster-level mediated effect, an applied researcher should first determine which of the mediated effects, or both, is important in answering the research questions being asked. Estimating a cross-level and cluster-level mediation process may provide very valuable information about the effect that an aggregated mediator has on an aggregated distal outcome, along with the impact that a mediator at the individual level has on a distal outcome at the individual level. In such a case, a very large number of clusters may be needed. However if only the cross-level mediated effect is of importance, then a smaller number of clusters may be needed to ensure reasonable parameter estimation. As was found in this study, conditions with 40 clusters resulted in generally well recovered estimates of the cross-level mediated effect. However, when fewer clusters are used, especially with fewer measurement occasions, convergence rates were found to be low. Applied researchers should also consider this issue to ensure that their data are sufficient to allow for convergence and a properly estimated model. Lastly, if an applied researcher is considering only estimating a cluster-level mediated effect, there needs to be a strong justification supporting estimation of such a model. Pituch and Stapleton (2012) state that in a model that only includes a cluster-level mediated effect the resulting mediated effect is a linear combination of the cross-level and cluster-level indirect effects. Given two models, one with cluster-only and one with both indirect effects, the same total indirect effect is estimated across the models, and therefore, are equivalent. The authors suggest that there is no advantage to restricting the modeling of the mediation process to include only the cluster-level mediated effect.

Increasing each of the four design factors that were manipulated in the simulation study would lead to better relative parameter bias and power for the mediated effects. However, an applied researcher may not easily be able to increase all factors. For instance, it is of course possible to increase the number of measurement occasions or the number of clusters, however, this will require use of additional resources which may be limited. To increase the actual mediated effect size is less directly possible although a researcher may attempt to choose strongly related intervention, mediating and outcome variables although researchers cannot ever know the true value of these relationships. In addition, to improve the proportion of explained variance in the mediator, a researcher should always select use of more reliable measures.

Practical Implications

Longitudinal data modeling is becoming more important as researchers and policymakers observe how treatments take time to show their effects and investigate change not only between but also within individuals. In addition, longitudinal mediation analysis further extends the analysis of change by allowing researches to not only help assess whether a program is a success or not but also how a program is a success through evaluating potential underlying mediators. The ability to use the growth of the mediator as one of the predictors of the growth of the distal outcome could thus lead to more nuanced understanding of treatments' effectiveness. For example, in a study conducted by Cheong et al. (2003), a treatment program for athletics led to positive change in perceived importance of their team leaders, which, in turn, led to positive change in their nutrition behaviors over time. In another example in longitudinal mediation design, math intrinsic motivation and achievement were modeled as two inter-related growth parameters and shown to be indirectly related to adult educational attainment through an intervening variable, high school math course accomplishments (Gottfried, Marcoulides, Gottfried, & Oliver, 2013). In addition, some of the latent growth parameters were shown to affect math course accomplishments both directly and indirectly through other latent growth parameters. While research is so far limited in terms of using the growth of an intervening variable as a predictor in the growth of the dependent variable, especially in cases with three levels, it is reasonable to assume that if a mediator is affecting a distal outcome, and many studies have involved cross-sectional mediation, then growth in the mediator might also influence development in the distal outcome over time. Findings of the mediation of a treatment's effect on growth in a distal outcome by growth in possibly a proximal mediating variable can help to inform ways to ultimately improve interventions and their effectiveness. This parallel process model in the context of mediation was already possible at two levels using the SEM framework but had yet to be parameterized using the multilevel modeling framework or extended to three levels in any framework.

Using the SEM framework with a parallel process model in longitudinal mediation may be appropriate in certain circumstances; however, this framework does have its limitations. There are scenarios in which the multilevel modeling framework would be more appropriate. One of the advantages of using a multilevel modeling framework to estimate a parallel process model is its ability to more easily conceptualize non-hierarchical clustered data structures, like cross-classified or multiple membership data structures. With constant mobility becoming prevalent in today's environment, non-hierarchical clusters are becoming more common, especially with longitudinal data spanning long time periods (see, for example, Herbers, Reynolds & Chen, 2013). SEM software (Mplus) has recently included the ability to estimate the basic three-level cross-classified model, but estimation of more complex non-hierarchical models has yet to be implemented in SEM software. In addition, the multilevel modeling framework handles adding levels, such as schools, more intuitively for some researchers. Spybrook and Raudenbush (2009) revealed that many recent educational designs involve four levels of sampling.

In addition, in the current study heteroscedasticity was modeled across measurement occasions and outcome type (mediator and distal outcome). As was noted earlier, typically in multilevel modeling, level-1 variances are assumed to be the constant across measurement occasions. However, as demonstrated in the illustrative ECLS-K real data example K, the level-1 residuals' variance components were mostly dependent on the *Time* predictor. Restricting level-1 variance components to be constrained to the same value across measurement occasions could result in loss of valuable information, especially when the proportion of explained variance change is helpful in evaluating a treatment's effectiveness over time.

Limitations

The model presented in this study can be used to assess longitudinal mediation for clustered participants such as for students in longitudinal CRTs. While this model is a useful extension of longitudinal mediation within the HLM framework, there are a number of limitations associated with this study including the restricted set of conditions that were examined in the simulation and analysis of data from a single real dataset. For instance, the proportion of explained variance to total variance was held constant across measurement occasions. However as was shown in the illustrative example, the proportion can reasonably be expected to change, and in most cases, might increase over time as the treatment variable takes effect. Further research should assess model estimation when this proportion is not assumed constant across time measurements. Another limitation was that the model that was estimated using real data included no additional predictors. Additional predictors are easily added to the model to investigate whether they explain some of the variability that was found. Also increasing the number of clusters to 100 in the pilot simulation study was found to improve bias substantially in the mediated effect estimates, especially in the cluster-level effect; however, bias was still found to be strongly negative and unacceptable for some conditions. Future research should examine various scenarios, such as further increasing the number of clusters or using different model specifications, which may improve bias more and result in better estimates for both mediated effects.

The simulated dataset was restricted to 300 converged replications per condition due to the long estimation completion time needed for each replication dataset. While the pilot study supported use of just 300 replications and only converged solutions were retained, clearly more replications would be needed for simulation studies investigating more complex models with larger datasets to ensure stable results. Future research should also consider different parameterizations of the model that may decrease its estimation completion time and also analyze the impact of the alternatives on the recovery of important parameters. For instance, an under-parameterized version of the model which has homogeneous level-1 variances and fixed covariance matrix elements could be assumed. Alternatively, an ad hoc two-step procedure that first estimates the mediator growth model and then uses the resulting predicted values of the growth trajectories at the individual and cluster levels as variables included in a second model of the distal outcome trajectory. Separating the two models may result in more quickly converged solutions. While information will be lost by constraining some parameters or by the ad hoc two-step procedure (when relevant covariances cannot be estimated), and the same problem in that the resulting mediated effect product term's distributional shape will be unknown, it is possible that the indirect and direct effects may still be recovered sufficiently well through use of MacKinnon and Lockwood's asymmetric confidence limits approach (2001). Future research could empirically assess these alternatives and compare results with those of the more parameterized model detailed in this study.

The ability to estimate these complex hierarchical models using MCMC estimation as well as the potential to manipulate resulting posterior distributions to directly test hypotheses about functions of parameters makes use of MCMC estimation particularly beneficial. However this direct assessment comes at the expense of the estimation completion time. Future research could also assess using other estimation methods like likelihood-based methods. However, again, the problem of the distribution of the resulting mediated effect would remain and have to be addressed and tested using bootstrapping or the empirical-M test.

In addition to introducing estimation of the three-level longitudinal mediation model, this study was also intended as a first assessment of this particular parameterization that can ultimately provide the foundation for additional extensions. Future methodological research should extend this model so that it can handle cross-classified and multiple-membership structures as well as additional functional forms rather than just linear. In addition, future research should examine estimation of an extension that allows the modeling of measurement error to extend to the second-order latent growth model (Hancock, Kuo, & Lawrence, 2001) to allow estimation of a parallel-process second-order growth model for clustered participants' data. This future research can help identify sample size and design guidelines supporting when these more complex multilevel latent variable regression models should be used. This research could also offer guidelines for cutoffs that indicate when model parameters are well recovered to provide more valid inferences about participants' growth trajectories and how these trajectories might mediate growth in parallel or more distal outcomes. Furthermore, more evaluation can be done that focuses on evaluating the recovery of variances and covariances at level-two and above.

Moreover, the study assumed that the parameters that represents the effect of the mediator's growth on the outcome's growth (b_B and b_W) and the treatment effect on both the mediator's and outcome's growth (a_B and c_B', respectively) as shown in Equation 32 are constant across the measurement periods. Situations may exist where the effects of either the independent variable or mediator change over time. In these situations, the cross-lagged model (Cole &

Maxwell, 2003) may be more applicable. Most of the variances for the mediator and outcome at each of the two levels (student and cluster) were assumed to be the constant. It is also possible that there are heterogeneous variances across students and clusters, and different variances should be taken into account. Further research should examine heterogeneous variances in a parallel process model under the multilevel framework. Also the independent variable was modeled to be dichotomous to simulate a treatment versus control group scenario, which occurs in many intervention studies. Further research could assess the effect of growth of an independent variable on growth of an outcome variable via growth of a mediating variable.

In addition, in the current study, the growth of the mediator is the only growth factor used as a predictor. Additional trajectory factors might be included as predictors, for example, the mediator's intercept might also affect growth in the distal outcome. Also, multiple mediators could exist in a model, where the independent variable has an indirect effect on the outcome variable through two or more mediators and their own growth trajectory parameters. The number of growth parameters that could affect the growth of a single outcome variable can increase substantially. In addition, the model was estimated using Bayesian analysis, which requires distributions on priors to be specified. Non-informative priors were used. Sensitivity analysis that assesses the effects of the choice of prior distributions provides another rich area for future research.

Last, in this study certain casual inference assumptions as described by VanderWeele (2010) were assumed for mediation to be inferred given the longitudinal, multilevel data. The assumptions include no treatment effect exists which might confound the relationship between the mediator and the outcome and that there was no within-cluster or between-cluster interference. Future research should investigate and assess conditions under which and how well longitudinal mediation data might meet these assumptions and help researchers understand how these assumptions might be compromised. If violations do occur, a more general approach can then be developed to refine the study design to avoid future violations.

Nevertheless, this study provides a useful starting point for further model extensions that can better match the complexities of longitudinal mediation models and data. Future research should extend what is proposed here and continue to expand the repertoire of possible models available that better fit the complexities of the real data encountered in social and educational research.

References

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*, 37-47.

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.

Bayarri, M. J., & Berger, J. O. (1999). Quantifying surprise in the data and model verification [with discussion]. In J. M. Bernardo et al. (Eds.), *Bayesian statistics 6* (pp. 53–82). Oxford, UK: Oxford University Press.

Bayarri, M. J., & Berger, J.O. (2000). P-values for composite null models [with discussion]. *Journal of the American Statistical Association, 95,* 1127-1142.

Beerli, P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics, 22*(3), 341-345.

Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research, 45,* 661-701. doi: 10.1080/00273171.2010.498292

Blozis, S.A., & Cho, Y.I. (2008). Coding and Centering of Time in Latent Curve Models in the Presence of Interindividual Time Heterogeneity. *Structural Equation Modeling, 15(3),* 413-433.

Browne, W. J., & Draper, D. (2006) A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis, 1*(3), 473-514. doi: 10.1214/06-BA117

Browne, W. J., Draper, D., Goldstein, H. & Rasbash, J. (2002). Bayesian and Likelihood methods for fitting multilevel models with complex level 1 variation. *Computational Statistics and Data Analysis, 39*.

Burton, A., Altman, D.G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine. 25*, 4279-4292. doi: 10.1002/sim.2673

Celeux, G., El Anbari, M., Marin, J.-M., Robert, C.P. (2012). Regularization in regression: comparing Bayesian and frequentist methods in a poorly informative situation. *Bayesian Anal, 2*, 477-502.

Cheong, J. (2011). Accuracy of Estimates and Statistical Power for Testing Meditation in Latent Growth Curve Modeling. *Structural Equation Modeling*, *18,* 195-211. doi:10.1080/10705511.2011.557334

Cheong, J., MacKinnon, D.P., & Khoo, S. T. (2003). Investigation of Mediational Processes Using Parallel Processes Latent Growth Curve Modeling. *Structural Equation Modeling*, *10*(2), 238-262.

Choi, K., & Seltzer, M. (2010). Modeling Heterogeneity in Relationships Between Initial Status and Rates of Change: Treating Latent Variable Regression Coefficients in a Three-Level Hierarchical Model. *Journal of Educational and Behavioral Statistics, 35*(1), 54-91. doi:10.3102/1076998609337138

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 558-577.

Enders, C.K. & Tofighi, D. (2007). Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue. *American Psychological Association, 12*(2), 121-138. doi: 10.1037/1082-989X.12.2.121

Farrant, B.M. & Zubrick, S.R. (2012). Early vocabulary development: The importance of joint attention and parent-child book reading. *First Language. 32*(3), 343-364. dio: 10.1177/0142723711422626

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515-533.

Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A., & Rubin D.B. (2014). *Bayesian Data Analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect effects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 243-259). Washington, DC: American Psychological Association.

Goodman, L.A. (1960). On the exact variance of products. *Journal of the American Statistical Association, 55,* 708-713.

Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., & Oliver, P. A. (2013). Longitudinal Pathways From Math Intrinsic Motivations and Achievement to Math Course Accomplishments and Educational Attainment. *Journal of Research on Educational Effectiveness, 6,* 68-92.

Grady, M., & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivariate Behavioral Research, 45*, 393-419.

Hale, D. R., Patalay P., Fitzgerald-Yau N., Hargreaves D.S., Bond L., Görzig A., Wolpert M., Stansfeld S.A., Viner R.M. (2014). School-level variation in health outcomes in adolescence: analysis of three longitudinal studies in England. *Preventive Science*, *15*(4), 600-610.

Hedeker, D. (2004). An introduction to growth modeling. In D. Kaplan (Ed.), *Quantitative Methodology for the Social Sciences.* Thousand Oaks CA: Sage Publications.

Hedges, L.V., & Hedberg, E.C. (2007a). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87.

Hedges, L. V. & Hedberg, E.C. (2007b). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education, 22*(10).

Herbers, J.E., Reynolds, A.J., & Chen, C. (2013). School mobility and developmental outcomes in young adulthood. *Development and Psychopathology, 25*(2), 501-515.

Hoogland, J.J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research, 26*(3), 329-367.

Huang, B., Sivaganasen, S., Succop, P., & Goodman, E. (2004). Statistical assessment of mediational effects for logistic mediational models. *Statistics in Medicine, 23*, 2713–28.

Kreft, I., De Leeuw, J. (1998). *Introducing Multilevel Modeling*. London, UK: Sage Publications.

Krull, J.L. & MacKinnon,D.P. (2001) Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*(2), 249-277.

Liu, L., Yu, L., Kalavacharla, V., & Liu, Z. (2011). A Bayesian model for gene family evolution. *BMC Bioinformatics, 12*, 426. doi: 10.1186/1471-2105-12-426

Localio, A.R., Berlin, J.A., Have, T.R. (2006). Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for binary outcomes: Bias and coverage of frequentist and Bayesian methods. *Statistics in Medicine, 25*, 2720-2736. doi: 10.1002/sim.2428

MacKinnon, D.P. (2000). Contrasts in multiple mediator models. In J. Rose, L. Chassin, C.C. Presson, & S.J. Sherman (Eds), *Multivariate applications in substance use research: New methods for new questions*, (pp. 141–160). Mahwah, NJ: Erlbaum.

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis.* New York, NY: Taylor & Francis Group, LLC.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, *17*, 144-158.

MacKinnon, D.P., Fairchild, A.J., & Fritz, M.S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614.

MacKinnon D.P., & Lockwood C.M. (2001). Distribution of products tests for the mediated effect. Arizona State University, Tempe; 2001. Unpublished manuscript.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*(1), 83-104. doi: 10.1037/1082-989X.7.1.83

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*, 99–128.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*, 41-62.

Marsh, H. W., Wen, Z. L., & Hau, K. T. (2006). Structural Equation Models of Latent Interaction and Quadratic Effects In G. Hancock & R. O. Mueller (Eds)., A second course in Structural Equation Modeling (pp.225-265). Greenwich, CT: Information Age Publishing.

Meeker,W. Q., Cornwell, L.W.,&Aroian, L. A. (1981). *Selected tables in mathematical statistics, Vol. VII: The product of two normally distributed random variables.* Providence, RI: American Mathematical Society.

Mehta, P.D., & West, S.G. (2000). Putting the Individual Back into Individual Growth Curves. *Psychological Methods*, *5*(1), 23-43.

Muthen, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 4(2),* 371-402.Muthén, L. K., & Muthén, B. O. (1998-2013). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Muthen, L. K., & Muthen, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599-620.

Pituch, K. A., Murphy, D. L., & Tate, R. L. (2010). Three-level models for indirect effects in school- and class-randomized experiments in education. *Journal of Experimental Education*, *78*, 60-95.

Pituch, K. A. & Stapleton, L. M. (2008). The Performance of Methods to Test Upper-Level Mediation in the Presence of Nonnormal Data. *Multivariate Behavioral Research, 43*, 237-67.

Pituch, K. A., & Stapleton, L. M. (2012). Distinguishing between cross- and cluster-level mediation processes in the cluster randomized trial. *Sociological Methods and Research*, *41*, 630-670. doi:10.1177/0049124112460380

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,* 367-400.

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

Preacher, K. J. (2011). Multilevel SEM Strategies for Evaluating Mediation in Three-Level Data. *Multivariate Behavioral Research*, *46*(4), 691-731. doi: 10.1080/00273171.2011.589280

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News, 6,* 7-11.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879-891.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raudenbush, S.W. and Bryk, A.S. (2002) *Hierarchical Linear Models* (Second Edition). Thousand Oaks, CA: Sage Publications.

Reardon, S. (2007). *Thirteen ways of looking at the black-white test score gap* (Working paper). Stanford, CA: Stanford University.

Robins, J. M., van der Vaart, A., & Ventura, V. (2000). Asymptotic distribution of p values in composite null hypotheses. *Journal of the American Statistical Association, 95,* 1143–1156.

Rovine M., & Liu S. Structural equation modeling approaches to longitudinal data. In: Newsom JT, Jones RN, Hofer SM, editors. Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences. New York, NY: Routledge; 2012. pp. 243–270.

Seltzer, M., Choi, K., & Thum, Y.M. (2003). Examining Relationships between Where Students Start and How Rapidly They Progress: Using New Developments in Growth Modeling to Gain Insight into the Distribution of Achievement within Schools. *Educational Evaluation and Policy Analysis, 25*(3), 263-286. doi:10.3102/01623737025003263

Serlin, R. C. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods, 5,* 230–240.

Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York. NY: Oxford University Press.

Snijders, T., & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* Sage: Thousand Oaks, CA.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290–312). Washington, DC: American Sociological Association.

Soest, T. von, & Hagtvet, K. A. (2011). Mediation Analysis in a Latent Growth Curve Modeling Framework. *Structural Equation Modeling*, *18,* 289-314. doi:10.1080/10705511.2011.557344

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS: Windows Version of Bayesian inference using Gibbs sampling, version 1.4, User Manual. MRC Biostatistics Unit, Cambridge University.

Spybrook, J. (2007). Examining the Experimental and Statistical Power of Group Randomized Trials Funded by the Institution of Education Sciences. Doctoral dissertation, University of Michigan. *Dissertation Abstracts International,* 68(10).

Spybrook, J., & Raudenbush, S.W. (2009). An examination of the precision and technical accuracy of the first wave of group randomized trials funded by the Institute of Education Sciences. *Educational Evaluation and Policy Analysis, 31*(3), 298-318.

Stan Development Team. 2014. RStan: the R interface to Stan, Version 2.5. http://mc-stan.org/rstan.html.

Stan Development Team. 2014. *Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0.*

Stoel, R.D., Van den Wittenboer, G. & Hox J.J. (2003). Analyzing longitudinal data using multilevel regression and latent growth curve analysis. *Metodologia de las Ciencas del Comportamiento*, *5*, 21-42.

Thomas, A., O Hara, B., Ligges, U., & Sturtz, S. (2006). Making BUGS Open. *R News. 6*, 12-17.

VanderWeele, T. (2010). Direct and Indirect Effects for Neighborhood-Based Clustered and Longitudinal Data. *Sociological Methods & Research, 38*(4), 515-544.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian Mediation Analysis. *Psychological Methods*, *14*(4), 301-322.