The Dissertation Committee for Lester Leland Lockhart III Certifies that this is the approved version of the following dissertation:

Nonlinear Mediation in Clustered Data: A Nonlinear Multilevel Mediation Model

Committee:

S. Natasha Beretvas, Supervisor

Gary Borich

Keenan Pituch

Paul von Hippel

Tiffany Whittaker

Nonlinear Mediation in Clustered Data: A Nonlinear Multilevel Mediation Model

by

Lester Leland Lockhart III, B.S.Psy.; M. 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 2012

Abstract

Nonlinear Mediation in Clustered Data: A Nonlinear Multilevel Mediation Model

Lester Leland Lockhart III, Ph.D.

The University of Texas at Austin, 2012

Supervisor: S. Natasha Beretvas

Mediational analysis quantifies proposed causal mechanisms through which treatments act on outcomes. In the presence of clustered data, conventional multiple regression mediational methods break down, requiring the use of hierarchical linear modeling techniques. As an additional consideration, nonlinear relationships in multilevel mediation models require unique specifications that are ignored if modeled linearly. Improper specification of nonlinear relationships can lead to a consistently overestimated mediated effect. This has direct implications for inferences regarding intervention causality and efficacy. The current investigation examined a specific nonlinear multilevel mediation model parameterization to account for nonlinear relationships in clustered data. A simulation study was conducted to compare linear and nonlinear model specifications in the presence of truly nonlinear data. MacKinnon et al.'s (2007a) empirical-M based PRODCLIN method for estimating the confidence interval surrounding the instantaneous indirect effect was used to compare confidence interval coverage rates surrounding both

the linear and nonlinear models' estimates. Overall, the nonlinear model's estimates were less biased, more efficient, and produced higher coverage rates than the linear model specification. For conditions containing a true value of zero for the instantaneous indirect effect, bias, efficiency, and coverage rate values were similar for the linear and nonlinear estimators. For conditions with a non-zero value for the instantaneous indirect effect, both the linear and nonlinear models were substantially biased. However, the nonlinear model was always less biased and always produced higher coverage rates than the linear model. The nonlinear model was more efficient than the linear model for all but two design conditions.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Mediational analysis quantifies proposed causal mechanisms through which treatments act on outcomes. In its most general conceptualization, mediation exists when the effect of a causal agent, *T*, on an outcome, *Y*, acts through at least one intervening variable, *M* (Hayes & Preacher, 2010). The validity of the causal implications resulting from a mediational analysis is largely dependent on the nature of the experimental design, as causality itself is an epistemological issue requiring the fulfillment of certain experimental conditions (Bauer, Preacher, & Gil, 2006). More specifically, causality can only be inferred when the variables involved covary with each other, when spurious results have been eliminated, and when hypothesized causes chronologically precede their presumed effects (Frazier, Tix, & Barron, 2004). Additionally, secondary manipulation of the mediating variable, *M*, along with experimental manipulation of the primary treatment variable, *T*, greatly strengthens the ensuing causal inferences (Spencer, Zanna, & Fong, 2005). With these issues in mind, mediational methods provide researchers the necessary statistical tools to investigate mediated mechanisms. This holds direct implications for modifying and improving treatment processes, as mediated effects shed light on the underlying causal mechanisms by which treatments are effective (Judd & Kenny, 1981). As such, this form of analysis is especially relevant to the social and behavioral sciences as investigators frequently seek causal explanations for treatment efficacy (Baron & Kenny, 1986).

In a typical mediation analysis, a treatment, *T*, is hypothesized to affect an outcome, *Y*, both directly and indirectly through some mediator, *M*. The portion of the treatment effect that acts through the mediator is labeled the *mediated* or *indirect effect*. To estimate this, *Y* is first regressed on *T* using conventional Ordinary Least Squares (OLS) methods to estimate the *total effect* of the treatment on the outcome. The regression coefficient for *T* is labeled the *c* parameter in mediation analysis. Next, *M* is regressed on *T* to estimate the effect of the treatment on the mediator itself. The effect of the treatment on the mediator is denoted the *a* parameter. *Y* is then regressed on both *M* and *T* to estimate the effect of the mediator on the outcome while controlling for the effect of the treatment. In this final regression, the coefficient for *M* is labeled the *b* parameter, and the coefficient for *T* (i.e, the *direct effect* of *T*) is labeled the *c'* parameter. The mediated effect is then calculated in one of two ways. The direct effect of the treatment on the outcome accounting for the mediator (*c'*) may be subtracted from the total effect of the treatment on the outcome (*c*) to estimate the indirect effect of the treatment on the outcome. Alternatively, the *a* parameter may simply be multiplied by the *b* parameter to estimate the indirect effect. In conventional OLS mediation analysis with non-nested data structures, the difference [$(c - c')$] and product of coefficients (*ab*) methods are equivalent (MacKinnon, 2008).

In mediation analysis, nested data structures require the use of multilevel modeling to partition variance at the appropriate levels accurately. The level at which the treatment, mediator, and outcome reside determines the specification of the multilevel models used in the ensuing mediation analysis. Though a variety of nesting structures

and model specifications exist in the literature, the current investigation will focus on two-level data structures whereby the treatment, mediator, and outcome all reside at level-1 and all level-1 participants (e.g., students) are nested within some hierarchical unit (e.g., schools). In the mediation literature, this is referred to as a $1 \rightarrow 1 \rightarrow 1$, or *lower-level*, mediation model. In this scenario, hierarchical linear modeling is used to estimate the *a, b*, *c*, and *c'* parameters. Here, the *ab* and $(c - c')$ estimates of the indirect effect are no longer equivalent, though previous research suggests that this discrepancy is typically minimal, especially with larger sample sizes (Krull & MacKinnon, 2001). In most mediation analyses, the *ab* estimate of the indirect effect is preferred to the $(c - c')$ estimate (Krull & MacKinnon, 1999).

Conventional multilevel mediation assumes linear relationships between the treatment, mediator, and outcome variables. In the context of mediation analysis, nonlinear relationships require unique model specifications. These may arise due to the presence of dichotomous outcomes or to the hypothesized relationships between the treatment, mediator, and outcome variables. Such nonlinear relationships often arise in the context of evaluating the functional form associated with hypothesized behavioral constructs (Hayes & Preacher, 2010). For example, the relationship between class size and subsequent student achievement has received considerable attention in recent educational research. Previous studies indicate that smaller class sizes may provide positive benefits for subsequent student achievement (Finn & Achilles 1990; Konstantopoulos 2008; Krueger 1999; Shin & Raudenbush, 2011). Given the variation in class size effects across schools (Konstantopoulos, 2011), multilevel methods are ideally

suited for handling such analyses. Mechanisms by which classroom interventions succeed, then, may provide differential effects based on the number of students contained in each class. In this scenario, class size may act as a treatment, functioning through some mediating variable to affect subsequent student achievement (as measured by scores on standardized exams or as binary indicators associated with matriculation to the subsequent grade level). Classroom size (as measured by the number of students enrolled in a particular class), however, may not be linearly related to a hypothesized mediator. More specifically, classroom sizes in excess of a specific number may not be linearly related to a mediator but may instead exhibit a ceiling effect beyond which all classroom sizes similarly affect the mediator. Large, lecture style college classes may provide one example of the ceiling effect of classroom size on mediated achievement effects. At the opposite end of the spectrum, small classroom sizes below some minimum threshold may exhibit similar floor effects on hypothesized mediators. Seminar and discussion based classes may exemplify this phenomenon.

As a starting point for extending nonlinear mediation for use in clustered data, the current investigation will examine a nonlinear treatment-mediator relationship in the presence of a linear mediator-outcome relationship. In the simplest, generic scenario, this relationship may arise because of the relationship between a dosage-like treatment variable and a mediator containing floor and ceiling effects. This nonlinear nature of this relationship is not captured when modeled using the conventional multilevel mediation paradigm. The current investigation provides an initial examination of one specific nonlinear model for estimating indirect effects in the presence of clustered data. As a

first attempt at examining this specific nonlinear parameterization, the current investigation will extend the current mediation literature by combining a specific multilevel mediation analysis design with a specific nonlinear functional form for the treatment-mediator relationship.

Chapter 2: Literature Review

The current study adds to the mediation literature by proposing an alternative parameterization to traditional mediation models, combining multilevel mediation analysis with a specific nonlinear *T*-*M* model specification. First, conventional linear mediation models are presented. Next, these traditional single-level models are extended to address hierarchically nested data via mixed linear models. A thorough examination of indirect effect calculation and hypothesis testing ensues, as this computation lies at the heart of all mediational analyses. This is followed by a discussion of nonlinear multilevel extensions for non-normally distributed data. At this point, problems with measures of mediation are discussed, followed by presentation of a nonlinear multilevel mediation model (NMMM) designed to address these concerns. A simulation study is then proposed to examine the accuracy of two specific indirect effect calculations as applied to the NMMM of interest.

Traditional Mediation

Standard mediational analyses quantify the causal relationships between an independent variable (or treatment, *T*), an outcome variable (or criterion, *Y*), and a mediating variable (or mechanism, *M*). In conventional mediation, the treatment variable typically consists of an indicator variable coded one for those in the treatment condition and zero for those in the control group. Alternatively, the treatment variable may consist of a continuous variable corresponding to a treatment dosage such as the amount in milligrams of a particular drug. Mediating variables are often operationalized as

constructs or measures hypothesized to describe how or why certain effects occur (Baron & Kenny, 1986; Bruner, 1957). Conceptually, mediation exists when a significant portion of the treatment effect operates through a mediating variable. Using path notation (Klein, 2005), Figure 1 depicts the total effect of *T* on outcome *Y*. However, part of this effect may function through an intermediate variable, *M*. The effect of *T* on *Y* is then broken down into its constituent paths as presented in Figure 2. Here, *a* represents the effect of *T* on *M*, *b* represents the effect of *M* on *Y*, and *c′* represents the effect of *T* on *Y* that is independent of *M*. *T* is hypothesized to affect *Y* both directly (path *c'*) and indirectly through *M* (paths *a* and *b*). The portion of this effect that goes through *M* is coined the *indirect effect*, whose calculation lies at the heart of mediational analysis.

Figure 1. Path Model Depicting Total effect, c, of Treatment T on Outcome Y.

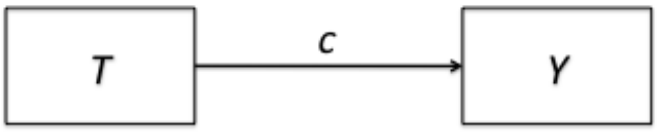
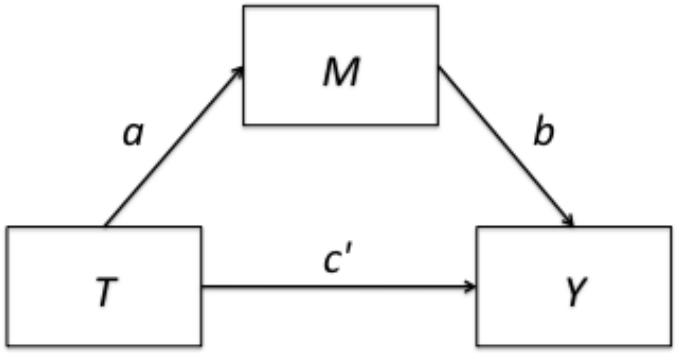


Figure 2. Single-level Linear Mediation Model of Treatment T's effect on Outcome Y through Mediator M.



In the simplest single mediator model, paths *a*, *b*, *c*, and *c′* in Figures 1 and 2 can be computed as unstandardized ordinary least squares (OLS) regression coefficient estimates obtained from the estimation of three regression models (MacKinnon, 2008). First, the total effect of a treatment variable (*T*) on an outcome variable (*Y*) is defined as the *c* parameter (Figure 1) by using the following linear equation:

$$Y_i + \beta_{(Y.T)0} + cT_i + e_{(Y.T)i}, \quad (1)$$

where Y_i represents the outcome score for subject *i*, $\beta_{(Y.T)0}$ represents the intercept for the prediction of *Y*, T_i represents the value of the treatment variable for subject *i*, *c* represents the total effect of the treatment on the outcome, and e_i represents the error term for subject *i*. This model corresponds to the path diagram in Figure 1, representing an estimate of the total effect of *T* on *Y* without accounting for possible covariates or mediators.

To obtain an estimate of path *a* in Figure 2, *M* is defined as a function of *T* according to the following model:

$$M_i = \beta_{(M.T)0} + aT_i + e_{(M.T)i}, \quad (2)$$

where M_i represents the score on the mediator for subject *i,* $\beta_{(M.T)0}$ represents the intercept for the prediction of *M*, T_i represents the treatment variable for subject *i* as in Equation 1, and $e_{(M.T)i}$ represents the error term for subject *i*. A third regression equation is used to provide estimates of *b* and *c′*. Here, *Y* is defined as a function of both *M* and *T*:

$$Y_i = \beta_{(Y.MT)0} + bM_i + c'T_i + e_{(Y.MT)i}. \quad (3)$$

In this equation, $\beta_{(Y.MT)0}$ represents the OLS intercept estimate for the prediction of *Y* from *M* and *T*, *b* represents the effect of the mediator on the outcome controlling for the effect of the treatment, *c′* represents the direct effect of the treatment on the outcome controlling for the mediator, and $e_{(Y.MT)i}$ represents the error term for subject *i*.

Once the path values are estimated, two primary methods exist for quantifying the indirect effect of *T* on *Y* through *M* (see, for example, MacKinnon, 2008). First, the indirect effect can be calculated according to the mathematical definition of the instantaneous indirect effect, θ (Stolzenberg, 1980):

$$\theta = \left(\frac{\partial M}{\partial T}\right)\left(\frac{\partial Y}{\partial M}\right). \tag{4}$$

As exposited by Hayes and Preacher (2010), the first partial derivative of a function with respect to a variable identifies the instantaneous rate of change of the former function (here, *M*) with respect to the second variable (here, *T*). In the context of multiple regression, this instantaneous rate of change is frequently interpreted as the effect of an independent variable on a criterion. Extending this to mediational analysis, the effect of *T* on *M* can be conceptualized as the rate of change of *M* with respect to *T*, $\left(\frac{\partial M}{\partial T}\right)$. Applying this to Equation 2,

$$\left(\frac{\partial M}{\partial T}\right) = a\,. \tag{5}$$

Similarly, the effect of *M* on *Y* can be conceptualized as the rate of change of *Y* with respect to *M* (controlling for *T*) [i.e., $\left(\frac{\partial Y}{\partial M}\right)$]. Applying this to Equation 3,

$$\left(\frac{\partial Y}{\partial M}\right) = b \ . \tag{6}$$

Assuming linear relationships between *T* and *M* and between *M* and *Y*, a one-point increase in *T* results in an *a*-point increase in *M*. In turn, a one-point increase in *M* results in a *b*-point increase in *Y*, holding *T* constant. Within this framework, the mathematical formulation of θ (Equation 4) defines the indirect effect for linear single mediator models as the product of the *a* and *b* parameter estimates as follows:

$$\theta = \left(\frac{\partial M}{\partial T}\right)\left(\frac{\partial Y}{\partial M}\right) = ab \, . \tag{7}$$

More intuitively, path tracing rules (see, for example, Kline, 2005) applied to Figure 2 provide the same results.

Alternatively, the indirect effect may be defined as the difference between the total effect of *T* on *Y* and the direct effect of *T* on *Y* accounting for *M*. Mathematically, this conceptualization of the indirect effect is calculated as $(c - c')$, where *c* and *c′* stem from Equations 1 and 3, respectively. In standard single mediator model analyses, the *ab* and $(c - c')$ methods are equivalent under OLS regression (MacKinnon, 2008).

Use of OLS regression for estimating the various paths depicted in Figures 1 and 2 (see Equations 1 through 3) works well for single-level datasets with no clustering of level-1 units within higher-level units. However, treatments, mediators, and outcomes are

frequently nested within various hierarchical units as a result of both group-level interventions and the natural clustering often inherent in organizational settings. Mediational analyses involving these types of data require proper parameterization via hierarchical linear modeling techniques to correctly partition variance at the appropriate levels. The following section describes this use of multilevel modeling to investigate mediation with clustered data.

MULTILEVEL MEDIATION

Clustered data presents problems for conventional mediational analysis as described in the previous section. More specifically, the dependence of observations within a particular cluster results in no partitioning of the constituent variance components if the clustering is ignored. In standard single-level multiple regression, correlated errors, as induced by the nesting of individuals within a level-2 unit, lead directly to underestimated standard errors, overestimated test statistics, and inflated type I error rates (Raudenbush & Bryk, 2002; Barcikowski, 1981; Moulton, 1986; Scariano & Davenport, 1987; Scott & Holt, 1982; Walsh, 1947). Failure to properly model all levels of nesting further complicates the inferences based on hypothesis tests (Moerbeek, 2004). The severity of the clustering effect can be assessed with the intraclass correlation (ICC), interpreted as the correlation between individuals belonging to the same level-2 unit:

$$ICC = \frac{MS_B - MS_W}{MS_B + (k-1)MS_w}, \tag{8}$$

where MS_B and MS_W represent the mean squared error between and within clusters, respectively, and k represents the number of subjects in each cluster (MacKinnon, 2008).

The ICC ranges from $\frac{-1}{(k-1)}$ to 1, with positive values indicating a violation of the independence assumption required in standard multiple regression analyses. The ICC is chi-square distributed, and as such can be tested for significance according to the following F-statistic (MacKinnon, 2008):

$$F_{g-1,g(k-1)} = \left(\frac{1+(k-1)ICC}{1-ICC} \right), \tag{9}$$

where g is the number of level-2 clusters and k is the number of subjects in each cluster. In reality, although the ICC may not be statistically significant for all clustered data structures, small values can still distort significance tests (Kreft, 1996; Muthén & Satorra, 1995). As such, non-zero ICC values require proper modeling via hierarchical linear modeling methods.

Proper mediation model parameterization of clustered observations depends largely on the level at which the treatment, mediator, and outcome variables are measured. For example, in the context of group-based treatments, cluster randomized trials assign clusters of individuals to treatment conditions but assess outcomes at the level of the participant. The $L_T \rightarrow L_M \rightarrow L_Y$ notation (where L_k represents the level at which variable k is measured) is commonly used to distinguish the different possible mediation model designs. For example, in a cluster randomized trial, if the hypothesized mediator is measured at the individual level, this design is referred to as an "upper-level" or "$2 \rightarrow 1 \rightarrow 1$" mediation design (Kenny, Kashy, & Bolger, 1998). Alternatively, the mediator in a cluster randomized trial may be measured at the cluster level resulting in a

"$2 \rightarrow 2 \rightarrow 1$" mediation design. These designs, when parameterized to account for clustering at level-2, produce fixed indirect effect estimates with the same statistical properties as those resulting from single-level designs.

For example, a typical $2 \rightarrow 1 \rightarrow 1$ cluster randomized trial may randomly assign level-2 units (e.g., treatment sites) to different treatments with outcomes and mediators measured at the level of the individual subjects within each site. In this instance, estimates of *a*, *b*, *c*, and *c'* should be obtained using multilevel modeling. Using HLM notation modified from that of Raudenbush & Bryk (2002), the *c* parameter for the total effect of the treatment on the outcome is estimated by the following model for the outcome at level-1 for subject *i* in site *j*:

$$Y_{ij} = \beta_{(Y.T)0j} + r_{(Y.T)ij}, \tag{10}$$

and at level-2:

$$\beta_{(Y.T)0j} = \gamma_{(Y.T)00} + cT_j + u_{(Y.T)0j}, \tag{11}$$

where Y_{ij} represents the outcome for subject *i* in site *j*, $\gamma_{(Y.T)00}$ represents the overall intercept across all sites, *c* represents the total effect of the treatment on the outcome, T_j represents the level of treatment received in site *j*, and $r_{(Y.T)ij}$ and $u_{(Y.T)0j}$ represent level-1 and level-2 random effects, respectively. Here, clustering is accounted for by modeling $\beta_{(Y.T)0j}$ as randomly varying across level-2 units, indicated by the inclusion of the level-2 residual, $u_{(Y.T)0j}$. This model assumes that the level-1 residuals, $r_{(Y.T)ij}$, are independently and normally distributed with a mean of zero and a constant variance. Additionally, the

level-2 residuals, $u_{(Y.T)0j}$, are assumed independently and normally distributed with a mean of zero and a constant variance in addition to being independent of the level-1 residuals. To estimate the *a* parameter, the mediator model for subject *i* in site *j* at level-1 is specified as:

$$M_{ij} = \beta_{(M.T)0j} + r_{(M.T)ij}, \tag{12}$$

and at level-2:

$$\beta_{(M.T)0j} = \gamma_{(M.T)00} + aT_j + u_{(M.T)0j}, \tag{13}$$

where M_{ij} represents the score on the mediator for subject *i* in site *j*, $\gamma_{(M.T)00}$ represents the overall intercept across all sites, *a* represents the effect of the treatment on the mediator, T_j represents the treatment condition administered to cluster *j*, and $r_{(M.T)ij}$ and $u_{(M.T)0j}$ represent level-1 and -2 residuals, respectively. Again, level-1 residuals are assumed independently and normally distributed with a mean of zero and a constant variance. Level-2 residuals are assumed independently and normally distributed with a mean of zero and a constant variance in addition to being independent of the level-1 residuals. Similarly, *b* and *c′* estimates are obtained from a multilevel model for the outcome variable with the following equation at level-1:

$$Y_{ij} = \beta_{(Y.MT)0j} + bM_{ij} + r_{(Y.MT)ij}, \tag{14}$$

and at level-2:

$$\beta_{(Y.MT)0j} = \gamma_{(Y.MT)00} + c'T_j + u_{(Y.MT)0j}, \tag{15}$$

where Y_{ij} represents the outcome score for subject *i* in site *j*, *b* and *c′* correspond to the paths depicted in Figure 2, $\gamma_{(Y.MT)00}$ represents the overall intercept across all sites, and $r_{(Y.MT)ij}$ and $u_{(Y.MT)0j}$ represent level-1 and -2 residuals, respectively. Statistical assumptions for the level-1 and level-2 residuals are identical to those mentioned above for Equations 10 through 13. Although this specification produces a fixed *b* parameter, an alternative specification may allow this parameter to vary across level-2 units. The estimate of *a* obtained in Equation 13 is multiplied by the estimate of *b* in Equation 14 to obtain the indirect effect, which can then be tested for statistical significance (MacKinnon, 2008). Alternatively, the *c'* parameter obtained in Equation 15 can be subtracted from the *c* parameter obtained in Equation 11 to yield a $(c-c')$ estimate of the indirect effect. Although the *ab* and $(c-c')$ methods for calculating the indirect effect are not mathematically equivalent in multilevel mediation models, the difference in the indirect effect calculated by the *ab* method and the $(c-c')$ method is typically negligible (MacKinnon, 2008).

As an alternative to the $2\rightarrow1\rightarrow1$ design, treatment sites (i.e., level-2 units) may be randomly assigned to treatment conditions, with a mediator measured at the site level and the outcome measured at the individual level. This results in a $2\rightarrow2\rightarrow1$ design, requiring estimation of a single-level and multilevel model to provide the relevant parameter estimates. First, the total effect, *c*, of the treatment on the outcome is specified as in the $2\rightarrow1\rightarrow1$ design (Equations 10 and 11). Next, because the mediator and the treatment variables are measured at the same level (level-2), OLS regression is used to

estimate the effect of the treatment on the mediator (*a*). This is modeled as in Equation 2. Finally, since the outcome, *Y*, is nested within level-2 units, a multilevel model is used to model this clustering in estimating the *b* and *c* parameters. At level-1, the model for the outcome for subject *i* in level-2 unit *j* is:

$$Y_{ij} = \beta_{(Y.MT)0j} + r_{(Y.MT)ij}, \tag{16}$$

and at level-2:

$$\beta_{(Y.MT)0j} = \gamma_{(Y.MT)00} + bM_j + c'T_j + u_{(Y.MT)0j} \tag{17}$$

where $\gamma_{(Y.MT)00}$ represents the overall intercept, and all other parameters are defined as above. As before, this model assumes that the level-1 residuals are independently and normally distributed with a mean of zero and a constant variance. Additionally, the level-2 residuals are assumed independently and normally distributed with a mean of zero and a constant variance in addition to being independent of the level-1 residuals. As with $2 \rightarrow 1 \rightarrow 1$ designs, *ab* or $(c - c')$ estimates of the indirect effect can be calculated from their respective parameter estimates. These multilevel specifications of $2 \rightarrow 1 \rightarrow 1$ and $2 \rightarrow 2 \rightarrow 1$ cluster randomized designs are not exhaustive of all cluster randomized data structures or model specifications. See Pituch and Stapleton (2008) and Pituch, Stapleton, and Kang (2006) for other exemplar multilevel model specifications.

In contrast to designs where sites or clusters are randomly assigned to treatment conditions, an intervention may be randomly assigned to individuals that are clustered within higher-level units. In these "$1 \rightarrow 1 \rightarrow 1$" or "lower-level" mediation designs (Kenny, Korchmaros, & Bolger, 2003), all variables included in the mediation analysis

are measured at the lowest level (level-1). These designs may contain either fixed or random indirect effect estimates that may require unique considerations. Given the clustering of individuals within relevant contexts (e.g., students within classrooms, patients within hospitals, etc.) multilevel modeling must be used to handle the resulting dependence of observations.

In lower-level mediation models, the total effect *c* is estimated with the following model at level-1:

$$Y_{ij} = \beta_{(Y.T)0j} + c_j T_{ij} + r_{(Y.T)ij}, \tag{18}$$

and at level-2:

$$\begin{cases} \beta_{(Y.T)0j} = \gamma_{(Y.T)00} + u_{(Y.T)0j} \\ c_j = c + u_{(Y.T)1j} \end{cases}, \tag{19}$$

where Y_{ij} represents the outcome for subject *i* in cluster *j*, $\gamma_{(Y.T)00}$ represents the overall intercept, T_{ij} represents the value on the treatment variable for subject *i* in cluster unit *j*, c_j represents the total effect of the treatment on the outcome for cluster *j*, $r_{(Y.T)ij}$ represents the level-1 residual, and $u_{(Y.T)0j}$ and $u_{(Y.T)1j}$ represent random effects corresponding to $\beta_{(Y.T)0j}$ and c_j, respectively. This model assumes that the level-1 residuals are independently and normally distributed with a mean of zero and a constant variance. The level-2 residuals are assumed independently and normally distributed with a mean of zero and a constant variance in addition to being independent of the level-1 residuals. In this instance with multiple level-2 residuals, however, the level-2 residuals

are allowed to covary with each other. The *a*, *b*, and *c′* parameters can be obtained by estimating the following multilevel model for the mediator, *M*, at level-1:

$$M_{ij} = \beta_{(M.T)0j} + a_j T_{ij} + r_{(M.T)ij}, \tag{20}$$

and at level-2:

$$\begin{cases} \beta_{(M.T)0j} = \gamma_{M0} + u_M \\ a_j = a + u_{(M.T)1j} \end{cases}, \tag{21}$$

and for the outcome, *Y*, at level-1:

$$Y_{ij} = \beta_{(Y.MT)0j} + b_j M_{ij} + c' T_{ij} + r_{(Y.MT)ij}, \tag{22}$$

and at level-2:

$$\begin{cases} \beta_{(Y.MT)0j} = \gamma_{Y0} + u_{Yj} \\ b_j = b + u_{(Y.MT)2j} \\ c'_j = c' + u_{(Y.MT)3j} \end{cases}, \tag{23}$$

where *a*, *b*, and *c'* represent the corresponding parameters used in the indirect effect calculation. The statistical assumptions for the residuals are identical to those for the residuals in Equations 18 and 19. The intercept terms $\beta_{(M.T)0j}$ and $\beta_{(Y.MT)0j}$ have similar random effects specifications at level-2, indicating that the intercept for each level-2 unit varies from the overall mean intercept. In Equations 21 and 23, the effects of the *a*, *b*, and *c'* parameters are modeled as varying across level-2 units, as indicated by the presence of the $u_{(M.T)1j}$ term in Equation 21 and the $u_{(Y.MT)2j}$ and $u_{(Y.MT)3j}$ terms in Equation 23. In this scenario, with random *a* and *b* parameters, however, the expected value of the *ab*

product is no longer the product of the individual fixed effects parameters. Instead, the expected value for the indirect effect is as follows (Goodman, 1960):

$$E(a_j b_j) = ab + \sigma_{ab}, \tag{24}$$

where a_j and b_j represent random variables, and σ_{ab} represents the covariance between them. This covariance term must be included for unbiased estimation of the indirect effect. However, if either the *a* or the *b* parameter is modeled as fixed, (i.e., if $u_{(M.T)1j}$ or $u_{(Y.MT)2j}$ are removed from the model in Equations 21 and 23), the indirect effect is now fixed with an expected value equal to the *ab* product as before (Kenny et al., 2003).

Bauer et al. (2006) outlined a method for dataset configuration and linear model parameterization that provides estimates of *a*, *b*, and σ_{ab} when calculating the indirect effect in a $1 \rightarrow 1 \rightarrow 1$ design. To do so, Bauer et al. combine the models for both the mediator (Equations 20 and 21) and the outcome (Equations 22 and 23) into a single equation by including a dummy coded variable for each portion of the model pertaining to each outcome. Combining these two models' equations to allow estimation of a single model requires stacking the dataset so that each subject's data is contained in two observations or rows. One observation contains variables relevant to the mediator model (Equations 20 and 21), and the other observation contains variables necessary for the outcome model (Equations 22 and 23). Each observation also contains two dummy-coded indicators, S_Y and S_M, coded such that S_Y equals one if the associated observation contains data for the outcome model and zero otherwise, and S_M equals one if that observation contains data for the mediator model and zero otherwise.

For $S_Y = 1$, a new outcome variable, *Z*, contains *Y* values, and for $S_M = 1$, $Z = M$. This results in the following model at level-1:

$$Z_{ij} = S_M(\beta_{(M.T)0j} + a_j T_{ij}) + S_Y(\beta_{(Y.MT)0j} + b_j M_{ij} + c'_j T_{ij}) + r_{Zij}, \quad (25)$$

and at level-2:

$$\begin{cases} \beta_{(M.T)0j} = \gamma_{M0} + u_{Mj} \\ a_j = a + u_{(M.T)1j} \\ \beta_{(Y.MT)0j} = \gamma_{Y0} + u_{Yj} \\ b_j = b + u_{(Y.MT)2j} \\ c'_j = c' + u_{(Y.MT)3j} \end{cases}, \quad (26)$$

producing the following combined equation:

$$Z_{ij} = S_M[(\gamma_{M0} + u_{Mj}) + (a + u_{(M.T)1j})T_{ij}] + S_Y[(\gamma_{Y0} + u_{Yj}) + (b + u_{(Y.MT)2j})M_{ij} + (c' + u_{(Y.MT)3j})T_{ij}] + r_{Zij}. \quad (27)$$

This specification simultaneously provides all parameters for both the mediator and outcome model specifications, including covariance estimates for all model parameters. The random effects in Equation 26 are assumed normally distributed with means equal to the average population effects, and a covariance structure that, at its most parameterized, can assume that each random effect covaries with every other random effect. Borrowing notation from Bauer et al. (2006), the following covariance structure can be assumed:

$$\begin{pmatrix} \beta_{(M.T)0j} \\ a_j \\ \beta_{(Y.MT)0j} \\ b_j \\ c'_j \end{pmatrix} \sim N\left[\begin{pmatrix} \gamma_{M0} \\ a \\ \gamma_{Y0} \\ b \\ c' \end{pmatrix}, \begin{pmatrix} \sigma^2_{\beta_{(M.T)0j}} & & & & \\ \sigma_{\beta_{(M.T)0j},a_j} & \sigma^2_{a_j} & & & \\ \sigma_{\beta_{(M.T)0j},\beta_{Yj}} & \sigma_{a_j,\beta_{(Y.MT)0j}} & \sigma^2_{\beta_{(Y.MT)0j}} & & \\ \sigma_{\beta_{(M.T)0j},b_j} & \sigma_{a_j,b_j} & \sigma_{\beta_{(Y.MT)0j},b_j} & \sigma^2_{b_j} & \\ \sigma_{\beta_{(M.T)0j},c'_j} & \sigma_{a_j,c'_j} & \sigma_{\beta_{(Y.MT)0j},c'_j} & \sigma_{b_j,c'_j} & \sigma^2_{c'_j} \end{pmatrix} \right]. \quad (28)$$

All level-2 random effects are assumed independent of the level-1 residuals. The covariance between the *a* and *b* estimates, σ_{ab}, can then be obtained from estimating this model. Alternatively, the $(c-c')$ method of calculating the indirect effect avoids the need to estimate the covariance between the *a* and *b* parameters. This is because the total effect, *c*, in a $1\rightarrow1\rightarrow1$ mediation model is defined according to the following equation (Kenny et al., 2003):

$$c = c' + ab + \sigma_{ab} \ . \tag{29}$$

Subtracting *c'* from both sides of the equation provides the following formula for $(c-c')$ in $1\rightarrow1\rightarrow1$ mediation:

$$(c-c') = ab + \sigma_{ab} \ . \tag{30}$$

Hence, in the $1\rightarrow1\rightarrow1$ scenario, the $(c-c')$ method for calculating the indirect effect provides an easier solution to the problem of estimating the indirect effect because it eliminates the need to estimate the covariances between any parameters. Again, although the *ab* and $(c-c')$ methods for calculating the indirect effect are not mathematically equivalent in multilevel mediation models, the difference is typically negligible (MacKinnon, 2008).

Multilevel model specifications for the aforementioned multilevel mediation designs are not without limitations. Most notably, Preacher, Zyphur, and Zhang (2010) noted that the conflation of within- and between- groups variance, as is often encountered in multilevel modeling contexts, must be addressed prior to any mention of inferences regarding the nature of the indirect effect. Commonly, this conflation of variances is

addressed by using group-mean centered predictors at level-2 (see Raudenbush & Bryk, 2002, for a complete discussion), and by reporting results at both the between-and within-groups levels (Zhang, Zyphur, and Preacher, 2009). However, even this solution may result in biased indirect effect estimates in circumstances with few level-1 units and a low ICC (Neuhaus & McCulloch, 2006). To address the conflated variance problem in this circumstance, a multilevel structural equation modeling (MSEM) framework may be preferable in certain situations as it allows for the explicit estimation of the level-2 error responsible for the conflated variance.

Previous research on the utility of MSEM in the context of mediation analysis with clustered data has focused primarily on mediation analyses with linear relationships between the constituent paths (Preacher et al., 2010). Truly nonlinear multilevel relationships may return spurious results if modeled using conventional linear methods (Bauer & Cai, 2009). Although additional research has examined the ability of SEM to parameterize nonlinear relationships (Hayes & Preacher, 2010), this work was restricted to single-level models. The current investigation extends the conventional mediation framework by combining both nested data structures with a specific nonlinear *T-M* relationship. As a first attempt at estimating this specific model, the MSEM framework will be eschewed in favor of an HLM framework as HLM analyses are still considered more accessible to the applied researcher (as evidenced by their preferred use over that of MSEM). MSEM approaches also require balanced data, a requirement that is often unfulfilled in many clinical studies (Mehta & Neale, 2005) Additionally, the use of an HLM design will allow future researchers to extend the model of interest to easily include

an optional third level or cross-classified data structures, extensions that, while possible, are still limited in the MSEM framework (Preacher, 2011).

Once the *a*, *b*, *c*, and *c'* parameters have been calculated as described above, the statistical significance of the indirect effect may be tested. The following section outlines commonly used methods of assessing the statistical significance of indirect effects, identifying the strengths and weaknesses of each test.

TESTS OF THE STATISTICAL SIGNIFICANCE OF THE INDIRECT EFFECT

Once the indirect effect is calculated, various procedures exist to assess its statistical significance. Of these, the causal steps approach outlined by Baron and Kenny (1986) is by far the most commonly used procedure (Preacher & Hayes, 2008). The causal steps procedure requires that parameters *c*, *a*, and *b* be sequentially statistically significant in order to infer mediation. Partial or complete mediation is then based on the significance of the *c'* parameter. This approach, although computationally simple, lacks statistical power and lowers the observed Type I error rate (MacKinnon, 2008; Pituch, Whittaker, & Stapleton, 2005). Additionally, some researchers (Collins, Graham, & Flaherty, 1998; Judd & Kenny, 1981; MacKinnon, Krull, & Lockwood, 2000; Shrout & Bolger, 2002) have suggested that a significant *c* parameter is not necessary for mediation to occur, calling into question the viability of the causal steps method.

To address this, the joint significance approach (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) has been suggested. This approach requires that both the *a* and *b* parameters be statistically significant in order to infer mediation. This procedure does not require statistical significance of the estimate of the *c* parameter. This approach

has been found to improve power when compared to the causal steps approach when used with single-level designs (MacKinnon et al., 2002), although use of this approach is often criticized for failing to provide a confidence interval for the indirect effect (e.g., Pituch et al., 2006). Additionally, simulation studies have found that this approach provides lower power in multilevel designs than other statistical tests of the indirect effect (Pituch et al., 2005; Pituch et al., 2006).

In an effort to provide confidence limits for the indirect effect, Sobel (1982, 1986) used Gaussian confidence limits through the calculation of a z-score based on the *ab* product and its standard error. However, the assumption of normality associated with use of a standard normal critical statistic is violated because the sampling distribution of estimates of the *ab* product is skewed, kurtotic, or both depending on the true value of the *ab* parameter (Springer & Thompson, 1966; Craig, 1936; Lomnicki, 1967). As such, z-score-based critical values and associated confidence intervals as used with Sobel's test are inappropriate. This is evidenced by simulation study results showing an asymmetry in the proportion of replications in which the true value falls to the left versus the right of the Sobel-calculated confidence interval estimates (Stone & Sobel, 1990; MacKinnon, Lockwood, & Williams, 2004).

To address these issues, bootstrap resampling methods (Shrout & Bolger, 2002) have been used to derive the empirical sampling distribution of the *ab* product estimate. Confidence intervals can be constructed using the resulting empirical distribution, with intervals excluding zero interpreted as evidence for mediation. Although several versions of bootstrapping exist, the bias-corrected parametric percentile bootstrap has been shown

to be the most accurate of the available methods for both single-level and multilevel designs (MacKinnon et al., 2004; Pituch et al., 2006). In multilevel designs, researchers are given two options when bootstrapping confidence intervals for the indirect effect: bootstrapping of residuals or bootstrapping of cases. In general, bootstrapping of residuals is preferred to bootstrapping of cases for two reasons (Pituch et al., 2006). First, multilevel models assume that values for the explanatory variables are fixed across samples; resampling cases with replacement would clearly change the distribution of the explanatory variables, violating this assumption. Second, resampling cases at any one level of a multilevel design may fail to reproduce the dependency present in the data or may produce inefficient parameter estimates. Pituch et al. (2006) outline the steps required to produce appropriate bias-corrected parametric percentile bootstrap estimates of the indirect effect. To summarize their steps for a $1 \rightarrow 1 \rightarrow 1$ design:

1. Sample residuals for as many level-1 and level-2 units (as correspond with the original sample) from a normal distribution with a mean of zero and variance equal to the estimated variance associated with that level.
2. Substitute the residuals and the original sample's observed values on the treatment variable into the equation for the mediator (Equations 20 and 21) to produce each case's value on the mediator variable.
3. Substitute values of the mediator obtained in Step 2 along with the resampled level-1 and two residuals and treatment variable values into the equations for the outcome, *Y* (Equations 22 and 23).

4. Rerun the original mediation analysis (Equations 20 through 23) on the bootstrapped data to obtain estimates of the *a*, *b*, and *c'* parameters for each bootstrapped sample.

5. Repeat steps 1 through 4 for each of the number, n_b, of bootstrapped samples required (typically 1,000).

6. Compute the *ab* product for each of the n_b replications. The values at the 2.5^{th} and 97.5^{th} percentile in this distribution serve as the lower and upper limits of a non-bias-corrected 95% confidence interval.

7. Calculate the z-score (z_0) in the standard normal distribution that corresponds to the percentile of the original *ab* estimate in the sample of 1,000 bootstrap estimates.

8. Calculate $2z_0 \pm 1.96$ and identify the percentiles in the standard normal distribution associated with the resulting upper and lower limits' values. The bias-corrected upper limit of the 95% confidence interval is the bootstrap estimate of *ab* that corresponds to the percentile equivalent of $2z_0 + 1.96$. Similarly, the bias-corrected lower confidence interval limit is the bootstrap estimate of *ab* that corresponds to the percentile equivalent of $2z_0 - 1.96$.

This method, while both accurate and powerful, is computationally intensive and complicated to use for most applied researchers.

As an alternative, the empirical M-test (MacKinnon, Fritz, Williams, & Lockwood, 2007a; Aroian, 1944) provides asymmetric confidence intervals for the *ab*

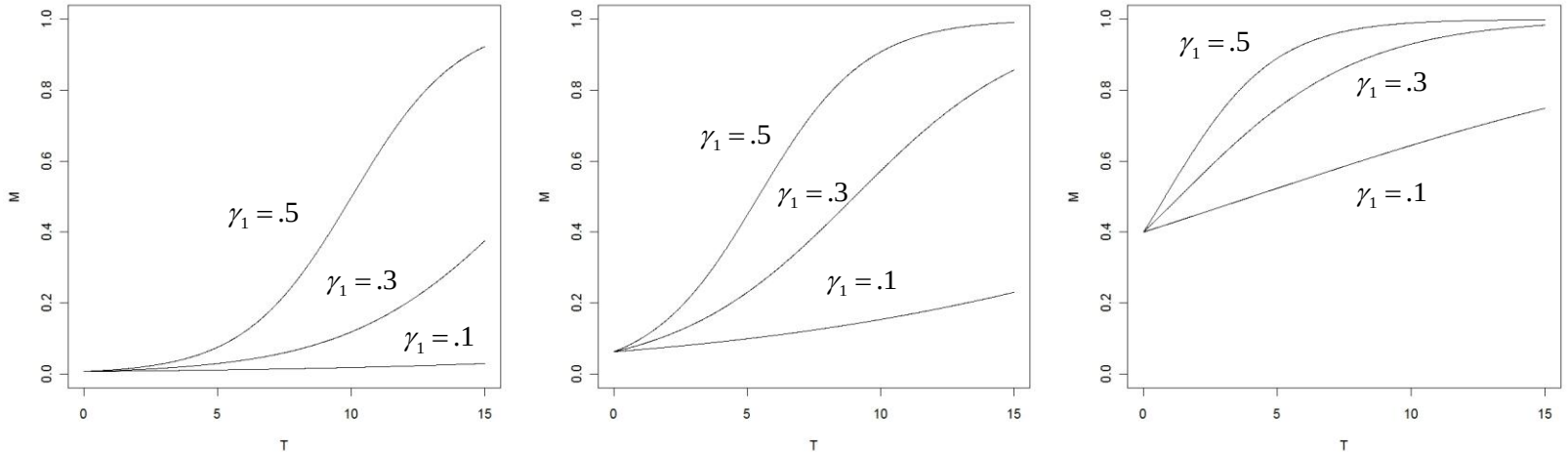
estimates by providing standardized critical values based on an approximation to the *ab* product's sampling distribution. This procedure performs similarly to bootstrap resampling methods, providing comparable statistical performance for single-level designs (MacKinnon et al., 2004), multilevel designs (Pituch et al., 2005; Pituch et al., 2006), and for non-normally distributed data (Pituch & Stapleton, 2008). To facilitate utilization of the empirical-M test, MacKinnon et al. (2007a) created the PRODCLIN program to compute asymmetric confidence intervals based on this approximation to the *ab* sampling distribution given a sample's estimates of *a* and *b* and their respective standard error values. Although bootstrap resampling methods provide slightly better statistical power, PRODCLIN's ease of use and overall accuracy render it preferable to all but bootstrap resampling procedures for linear models (Pituch et al., 2006).

The $(c - c')$ estimate of the indirect effect may also be tested for statistical significance by dividing this estimate by one of many analytically derived standard errors (see MacKinnon et al., 2002, for specific standard error formulations) and comparing the resulting value to a *t*- or *z*-distribution. As previously mentioned, the $(c - c')$ method is identical to the *ab* method under OLS regression but is slightly different from the *ab* estimate in multilevel analyses because of the discrepancy in the weighting matrices used to estimate the relevant fixed effects (Krull & MacKinnon, 2001). Although this difference is typically negligible, the *ab* estimator of the indirect effect is more efficient than the $(c - c')$ estimator in multilevel analyses. Additionally, the *ab* estimator provides information regarding specific indirect effects in multiple mediator models (Krull &

Mackinnon, 1999), whereas the $(c - c')$ estimator provides only an estimate of the total mediated effect (Krull & MacKinnon, 1999). Furthermore, all but bootstrap resampling methods for significance testing of the indirect effect estimate assume linear relationships between *T*, *M*, and *Y*. However, this assumption may be invalid. For example, the relationship between the exogenous variable, *T*, and the mediator, *M*, could be nonlinear. The mediator could be a discrete, dichotomous variable containing only values of zero or one, requiring a nonlinear specification to model the relationship between *T* and *M*. The particular function of interest in the current study refers to scenarios in which the relationship between *T* and *M* follows an ogive pattern (such as the functions depicted in Figure 3). This logistic ogive pattern can also occur when the interval-scaled mediator is a measure of a construct on which the scores exhibit a ceiling and/or floor effect.

Although the current investigation is specifically focused on modeling a nonlinear relationship between *T* and *M*, mediational analysis for nonlinear relationships is historically rooted in the development of path analytic methods for dichotomous outcomes (see Winship & Mare, 1983). As such, the following section outlines the issues associated with nonlinear mediation for dichotomous outcomes, as these are special cases of a more generalized nonlinear trajectory. This is followed by a discussion of a generalized approach to nonlinear mediation for mediating variables that are not necessarily dichotomous but with scores that exhibit floor and ceiling effects.

Figure 3: Nonlinear Parameters and their effects on the logistic change trajectory.



NONLINEAR MEDIATION FOR DICHOTOMOUS OUTCOMES

Dichotomous outcomes require special consideration in the context of statistical modeling. In many investigations, the outcome of interest, *Y*, consists of a dichotomous variable used to indicate the presence or absence of a specific condition. For example, in mediational analyses in the field of medicine, *Y* may represent a binary variable indicating the presence or absence of a specific disease in a patient after undergoing some intervention, *T*. In this situation, conventional linear analyses of the indirect effect are no longer appropriate for three primary reasons (MacKinnon, 2008). First, linear analyses produce predicted values outside of the range of possible values (of zero to one). For example, using standard multiple regression to predict a binary outcome can result in predicted values less than zero or greater than one for certain observed combinations of the independent variables. Second, the standard errors of the regression coefficients are inaccurate, complicating the interpretation of the ensuing inferential statistics. Finally, the residuals for models containing binary outcomes are not normally distributed, directly violating the assumption of normally distributed errors associated with linear regression model estimation techniques. Given these issues, standard linear models (and their multilevel extensions) are inappropriate in cases where the criterion of interest is binary.

Logistic regression addresses problems associated when estimating models with binary outcomes and is the method most commonly used to analyze dichotomous dependent variables (Hosmer & Lemeshow, 2000). This class of procedures provides upper and lower asymptotes of zero and one, respectively, which correspond to the maximum and minimum values of the observed dichotomous outcomes. These models

estimate the log-odds of success on the outcome variable according to the following specification:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n, \quad (31)$$

where p_i is the probability of success on the outcome variable for case *i*, $1-p_i$ is the probability of failure for case *i*, $\left(\frac{p_i}{1-p_i}\right)$ is the odds of success, X_1 through X_n are a set of *n* predictor variables, and β_0 to β_n are the parameter estimates associated with their respective independent variables. Solving for the probability of success (e.g., Howell, 2002):

$$p_i = \frac{1}{1+\exp[-(\beta_0 + \beta_1 X_1 + ... + \beta_n X_n)]}. \quad (32)$$

Previous research on nonlinear mediation analysis has focused primarily on scenarios with a dichotomous distal outcome variable, *Y* (Mackinnon, 2008). In these analyses, path *a* in Figure 2 can be estimated via conventional linear methods when it is assumed that the relationship between *T* and *M* is linear. However, paths *b*, *c*, and *c*' require the use of logistic techniques when *Y* is dichotomous. In this situation, binary outcomes are commonly modeled as having continuous underlying counterparts whose latent values are deterministically or stochastically related to the observed dichotomous outcome (Winship & Mare, 1983). Although several latent variable conceptualizations exist to model the relationship between the unobserved latent variable, *Y**, and the observed dichotomous outcome, *Y*, the threshold model as outlined by Winship and Mare

is commonly used to specify this relationship in meditational analysis (e.g., MacKinnon, D. P., Lockwood, C. M., Brown, C. H., Wang, W., & Hoffman, J. M., 2007b; MacKinnon, 2008). This model specifies that the observed dichotomous outcome, *Y*, and the unobserved latent continuous variable, *Y**, are related as follows:

$$\begin{cases} Y = 1 if\ Y^* \geq L \\ Y = 0 if\ Y^* < L \end{cases}, \tag{33}$$

where *Y** is assumed to have a mean of zero and a variance of one, and *L* is the threshold across which *Y* changes from zero to one. This implies that *Y* and *Y** are directly related through a nonlinear transformation such that all values of *Y** greater than or equal to *L* have been transformed to 1, and all values of *Y** below *L* have been transformed to zero (Winship & Mare, 1983). In turn, *Y** may be related to a set of observed continuous predictors, X_1,...., X_n. This relationship may be modeled linearly as:

$$Y^* = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon_{Y^*} \tag{34}$$

Here, β_0 through β_n are parameters to be estimated, and ε_{Y^*} is an error term that is assumed to be uncorrelated with all X_n. If ε_{Y^*} is assumed to follow an extreme value distribution (i.e., is considered a value in excess of a predetermined threshold; Johnson & Kotz, 1970) then Equations 33 and 34 above define a logit, or logistic regression, model as presented in Equations 31 and 32 (McFadden, 1974; Winship & Mare, 1983). Given this framework, logistic regression methods are equivalent to structural equation models whereby a dichotomous variable serves as an indicator of an unobserved latent variable (i.e., the natural log of the odds of subject *i* possessing a score of one on the outcome

variable). Alternatively, if ε_{Y^*} is assumed normally distributed, then Equations 33 and 34 define a probit, or probit regression, model (Winship & Mare, 1983; Hanushek & Jackson, 1977). The choice of the ε_{Y^*} distribution is somewhat arbitrary, as the logit and probit models are essentially interchangeable given the similarity between the logistic and cumulative normal distribution functions (Winship & Mare, 1983; Hanushek & Jackson, 1977). The current investigation will focus on the logistic regression conceptualization, with the understanding that probit regression provides an alternative and nearly equivalent mode of analysis.

The model specified within the threshold framework described by Equations 33 and 34 is underidentified because the scale of the unobserved latent variable *Y** is not directly observed (Winship & Mare, 1983). To address this, a scaling assumption is required regarding either the variance of *Y** or of ε_{Y^*}. Commonly, the standard logistic regression error, $\frac{\pi^2}{3}$, is substituted for ε_{Y^*} to identify the model. Winship and Mare provide a discussion of this issue. With the threshold model for a dichotomous outcome thus defined, both linear and logistic regression models are used to estimate the *a*, *b*, *c*, and *c'* parameters in a meditational analysis. First, the total effect is estimated from the logistic regression of the outcome, *Y*, on the treatment, *T*:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_{(Y^*.T)0} + cT_i \tag{35}$$

where p_i is the probability of subject *i* containing a 1 on the outcome variable, $\beta_{(Y^*.T)0}$ represents the intercept for the logistic equation, and all other variables are defined as before. Next, the *a* parameter is estimated linearly as in Equation 2. Finally, the *b* and *c'* parameters are estimated from the logistic regression of *Y* on *T* and *M*:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_{(Y^*.MT)0} + bM_i + c'T_i \tag{36}$$

where $\beta_{(Y^*.MT)0}$ is the intercept for the logistic regression, and all other parameters are defined as before. As mentioned above, the residual error in Equations 35 and 36 is fixed at $\frac{\pi^2}{3}$.

Once the scale of the residual variance in logistic regression is fixed at $\frac{\pi^2}{3}$, the scale for the latent continuous outcome variable *Y** can vary across models (MacKinnon & Dwyer, 1993; Jasti, Dudley, & Goldwater, 2008). Put differently, the variance of the dependent variable in logistic regression is fixed at $\frac{\pi^2}{3}$ regardless of the number of variables included as predictors. Consequently, because the model for the outcome (Equation 36) includes more predictors than the model for the total effect (Equation 35), estimates for the *b*, *c,* and *c*' parameters depend heavily on the other variables in the model, and the $(c-c')$ method of calculating the indirect effect is no longer equivalent or nearly equivalent to the *ab* method (MacKinnon & Dwyer, 1993). Under these conditions, simulation studies suggest that the *ab* estimate of the indirect effect is less

biased and more robust to assumption violations than is the $(c - c')$ estimate (MacKinnon, 2008; MacKinnon & Dwyer, 1993).

Additionally, MacKinnon and Dwyer (1993) suggest standardizing estimates of the *b* parameter and its standard error. In logistic regression, this is accomplished by dividing each parameter in the logistic model by the standard error for the logistic regression model (i.e., the square root of the residual error variance for the logistic regression model). As described by Winship and Mare (1983) and MacKinnon et. al. (2007b), the residual error variance, σ_{Y^*}, for a logistic regression mediation model is calculated as follows:

$$\sigma_{Y^*} = \sqrt{c'^2 \sigma_T^2 + b^2 \sigma_M^2 + 2c' b \sigma_{TM} + \frac{\pi^2}{3}}, \qquad (37)$$

where *c'* and *b* represent the parameter estimates from the logistic regression, σ_T^2 represents the variance of the treatment variable, σ_M^2 represents the variance of the mediator, σ_{TM}^2 represents the covariance between the treatment and the mediator, and $\frac{\pi^2}{3}$ is the fixed residual variance from the logistic regression. Once standardized, model coefficients now represent the effect of a unit change in a respective independent variable in standard deviations of the latent variable *Y** (Winship & Mare, 1983). For example, the estimate of the *b* parameter standardized by dividing by Equation 37 provides an estimate of the change (in standard deviation units) on the continuous latent variable *Y** that results from of a one unit change in the mediator, *M*, controlling for the treatment variable.

Once *b* and its standard error are appropriately standardized, three methods exist for quantifying the indirect effect. First, the product of *a* with standardized *b* may be used as an estimate of the mediated effect (MacKinnon & Dwyer, 1993), and the standard error of *a* and the standardized standard error of *b* can be used to test the statistical significance of the *ab* product. This is directly analogous to the use of the *ab* product as an estimate of the indirect effect assuming a linear relationship between all variables. However, when the outcome, *Y*, is dichotomous, the *ab* method of estimating the indirect effect systematically overestimates the true mediated effect (Li, Schneider, and Bennett, 2007). This is because the true value of the indirect effect in a logistic regression analysis is defined by the instantaneous indirect effect in Equation 4. As such, the second method for quantifying the indirect effect utilizes the mathematical definition of the instantaneous indirect effect (Equation 4). In this instance, where logistic regression is used to analyze a binary outcome *Y* and the relationship between *M* and *T* is modeled linearly as in Equation 2, Winship & Mare (1983) and Li et. al. (2007) derive the true indirect effect as the instantaneous indirect effect, θ. First, the partial derivative of *M* is taken with respect to *T*:

$$\frac{\partial M}{\partial T} = a. \tag{38}$$

Next, the partial derivative of *Y* is taken with respect to *M*:

$$\frac{\partial Y}{\partial M} = b\left\{\frac{\exp(\beta_{0(Y^*.MT)} + bM + c'T)}{[1+\exp(\beta_{0(Y^*.MT)} + bM + c'T)]^2}\right\}. \tag{39}$$

Finally, Equations 38 and 39 are multiplied to obtain the instantaneous indirect effect:

$$\theta = \left(\frac{\partial M}{\partial T}\right)\left(\frac{\partial Y}{\partial M}\right) = ab\left\{\frac{\exp(\beta_{(Y^*.MT)0} + bM + c'T)}{[1+\exp(\beta_{(Y^*.MT)0} + bM + c'T)]^2}\right\}. \quad (40)$$

Here, the true indirect effect varies as a function of both *M* and *T*. This provides a form of moderated mediation, whereby the true indirect effect depends on both the value of the mediator and the dosage of treatment received. Hayes and Preacher (2010) recommend examining the true indirect effect at the mean of *T* as well as at one standard deviation above and below the mean of *T* to investigate the instantaneous indirect effect at low, average, and high values of treatment dosage, although any conditional values of interest will suffice. Confidence intervals for the instantaneous indirect effect at each value of the treatment dosage can be estimated via the bias-corrected parametric percentile bootstrap resampling method described above. Additionally, although meditational analyses often include dummy coded treatment indicators, Equation 40 assumes that the treatment is measured continuously. As such, the current investigation is concerned only with continuous, dosage-like treatment variables, acknowledging that additional derivations are necessary for use with binary treatment indicators. See Li et al. (2007) for a discussion.

From the formulation of the instantaneous indirect effect in Equation 40, it is evident that the instantaneous indirect effect for continuous *T* is simply the product of the *a* and *b* coefficients adjusted by a factor that depends on the value of the mediator and the dosage of treatment received. Because this adjustment factor will always be less than one, the *ab* estimate of the indirect effect will always overstate the magnitude of the mediated effect (Li et al., 2007). The third method for calculating the indirect effect

utilizes the $(c-c')$ approach outlined above. However, in cases where the outcome, *Y*, is dichotomous, this method is generally more biased than the already biased *ab* method (MacKinnon et al., 2007b).

Mediating relationships involving binary mediators also require the use of nonlinear models to appropriately handle estimation of indirect effects (Li et al., 2007; Winship & Mare, 1983; Jasti et al., 2008). This form of analysis is particularly useful when the mediator indicates the presence or absence of a specific mediating condition. As in the case with dichotomous outcomes, an observed dichotomous mediator, *M*, is specified as a threshold model whereby *M* is modeled as an observed indicator of an unobserved latent variable, *M**. *M* and *M** are then related through a nonlinear transformation defined by:

$$\begin{cases} M = 1 if\ M^* \geq L \\ M = 0 if\ M^* < L \end{cases}, \tag{41}$$

where *M** is assumed to have a mean of zero and a variance of one, and *L* is the threshold across which *M* changes from zero to one. This model specification directly parallels the threshold model specification for dichotomous *Y* (see Equation 33). As before, M* may be related to a set of observed continuous predictors, $X_1,....,X_n$. This relationship may be modeled linearly as:

$$M^* = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon_{M^*} \tag{42}$$

Here, β_0 through β_n are parameters to be estimated, and ε_{M^*} is an error term that is assumed to be uncorrelated with all X_n. If ε_{M^*} is assumed to follow an extreme value

distribution (Johnson & Kotz, 1970) then Equations 41 and 42 define a logit, or logistic regression, model (McFadden, 1974; Winship & Mare, 1983), although probit models provide viable alternatives. As previously mentioned, the current investigation will focus on the logistic regression conceptualization.

In mediation analysis with a binary mediator, only the *a* parameter is estimated using logistic regression:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_{(M^*.T)0} + aT_i \tag{43}$$

where p_i is the probability of subject *i* having a score of one on the mediator, and all other variables are defined as above. The *c* parameter is modeled linearly and estimated according to Equation 1. Similarly the *b* and *c'* prime parameters are modeled according to Equation 3. These estimates are then used in calculating the indirect effect as previously defined.

As in the case of dichotomous *Y*, the variance for the mediator model must be fixed in order for the model to be identified. Again, the residual variance is conventionally fixed at $\frac{\pi^2}{3}$. Now, with a binary mediator instead of a binary outcome, the *a* path and its standard error must be standardized by dividing both by the standard error for the logistic regression model with dichotomous outcome *M* (Jasti et al., 2008):

$$\sigma_{M^*} = \sqrt{a^2\sigma_T^2 + \frac{\pi^2}{3}} \tag{44}$$

Equation 44 contains fewer terms than the standard error for dichotomous *Y* in Equation 37 because the logistic regression model of the mediator (Equation 43) contains only one predictor: the treatment, *T*. In contrast, the model for a dichotomous *Y* contains two predictors: the treatment, *T*, and the mediator, *M*. This explains the simplicity of the equation for the standard error of a dichotomous mediator (Equation 44) compared to the standard error for a dichotomous outcome (Equation 37).

Assuming a linear relationship between *M* and *Y* and between *T* and *Y*, the *b* and *c′* paths can be estimated using the traditional linear model, and the product of *b* and standardized *a* provide a measure of the indirect effect. However, as is the case with binary *Y*, the *ab* product is an overestimate of the true mediated effect (Li et al., 2007). Additionally, in the presence of a binary *M*, the $(c - c')$ method is sensitive to the skew of the distribution of the treatment dosage, and as such should be used cautiously (Li et al., 2007). The true indirect effect is again defined as the instantaneous indirect effect described in Equation 4. With logistic regression used to analyze the relationship between a binary mediator and the treatment variable, and all other relationships modeled linearly, Winship & Mare (1983) and Li et. al. derive the true indirect effect as the instantaneous indirect effect, θ. First, the derivative of *M* is taken with respect to *T*:

$$\frac{\partial M}{\partial T} = a\left\{\frac{\exp(\beta_{(M^*.T)0} + aT)}{[1+\exp(\beta_{(M^*.T)0} + aT)]^2}\right\}. \tag{45}$$

Next, the derivative of *Y* is taken with respect to *M*:

$$\frac{\partial Y}{\partial M} = b. \tag{46}$$

Finally, Equations 45 and 46 are multiplied to obtain the instantaneous indirect effect:

$$\theta = \left(\frac{\partial M}{\partial T}\right)\left(\frac{\partial Y}{\partial M}\right) = ab\left\{\frac{\exp(\beta_{(M^*.T)0} + aT)}{[1+\exp(\beta_{(M^*.T)0} + aT`)]^2}\right\}. \quad (47)$$

Here, the true indirect effect varies as a function of *T*. This, again, provides a form of moderated mediation, whereby the true indirect effect depends on the level of treatment received. Equation 47 assumes that the treatment is measured continuously. As before, the instantaneous indirect effect for continuous *T* is simply the product of the *a* and *b* coefficients adjusted by a factor that depends on the dosage of treatment received. Since this adjustment factor will always be less than one, the *ab* estimate of the indirect effect will always overestimate the magnitude of the mediated effect (Li et al., 2007). As is the case for binary *Y*, the instantaneous indirect effect may be evaluated at the mean of *T* and at one standard deviation above and below the mean of *T* to investigate the value of the instantaneous indirect effect at low, average, and high values of the treatment dosage (Hayes & Preacher, 2010). Additionally, as previously mentioned, the bias-corrected parametric percentile bootstrap resampling method may be used to provide confidence intervals around the instantaneous indirect effect at each dosage of the treatment.

In summary, binary outcomes in mediational analyses require special considerations in estimating and in testing the statistical significance of the indirect effect. When properly formulated, this indirect effect depends on the dosage of treatment received, providing a form of moderated mediation. As an additional consideration, many constructs in the social and behavioral sciences contain properties similar to binary outcomes, particularly with regard to survey scores used as measures of underlying

mediators. These surveys frequently contain minimum and maximum possible scores (termed floor and ceiling effects, respectively) resulting in data with nonlinear relationships mimicking the pattern observed with binary outcomes. These constructs also require special treatment to properly model their lower and upper asymptotes and the resulting nonlinear relationships found between the score on the mediator and other variables included in the model. The following section addresses these concerns associated with measures of mediation frequently used in the social and behavioral sciences.

NONLINEAR MEDIATION FOR CONTINUOUS MEDIATORS WITH FLOOR AND CEILING EFFECTS

Hypothesized models in the social and behavioral sciences often posit nonlinear relationships amongst variables. For example, as cited in Singer and Willett (2003), Robertson (1909) theorized that the rate at which learning occurs is proportional to the amount of learning that has previously occurred times the amount of learning that will occur in the future. Mathematically, this can be expressed as a differential equation of the form:

$$\frac{dY}{dt} = kY(\alpha - Y), \tag{48}$$

where $\frac{dY}{dt}$ is the rate of learning, Y is the amount learned by time t, α is an upper limit to the amount that can be learned, and k is a proportionality constant. This first-order differential equation has a solution of the form:

$$Y_i = \frac{\alpha_i}{1 + \pi_{0i} \exp\left[-(\pi_{1i} TIME_i)\right]} \quad , \tag{49}$$

which relates learning and time through an exponential (nonlinear) function for subject *i*. These functions are similar to the logistic function used in the analysis of dichotomous outcomes (see Equation 32). However, the function in Equation 49 differs from the logistic function in two important ways. First, Equation 49 allows for an upper asymptote value other than one, indicated by the α_i in the numerator. Second, the π_{0i} coefficient allows for specification of a model whereby the function crosses the Y-axis at a non-zero value. Despite these differences between the logistic function and the function described in Equation 49, both stem from the same family of nonlinear distributions.

Nonlinear multilevel model specifications are not limited to the logistic trajectory exposited above. Polynomial specifications (such as quadratic and cubic parameterizations) and myriad other exponential relationships provide viable alternatives to the nonlinear model described in Equation 49 (see Bauer & Cai, 2009, Cudeck & Harring, 2007, Hayes & Preacher, 2010, Preacher, Zyphur, & Zhang, 2010, and Singer & Willett, 2003, for other exemplar nonlinear model specifications). In the context of measures with floor and ceiling effects, polynomial models are of limited utility: polynomial functions do not "flatten-out" asymptotically but, rather, continue infinitely in both the positive and negative directions. Using polynomial models on measures with floor and ceiling effects is akin to using linear regression on data with dichotomous outcomes; although this is computationally feasible, linear regression estimates with binary outcomes provide predicted values outside the range of possible outcomes.

Additionally, the parameters associated with polynomial functions are often difficult to interpret (Cudeck & du Toit, 2002). These considerations, combined with previous research regarding the use of Equation 49 with clustered data, suggest that this model specification may provide a viable option in the context of multilevel mediation modeling. As such, the current study will focus on the nonlinear specification in Equation 49.

Nonlinear relationships abound in the social and behavioral sciences (for examples, see Debreu, 1959; Yerkes & Dodson, 1908; Kahneman & Tversky, 1979; Knobloch, 2007). Thus, it is conceivable that mediation analyses might involve a nonlinear relationship between *T* and *M* stemming from use of a mediation measure, for example, that contains floor and ceiling effects. This is often the case when measures of psychological constructs are used as a mediating variable. For example, scores on criterion-referenced tests commonly include both floor and ceiling effects that should be accounted for when modeling the mediating effect. If the effect of *T* on *M* is modeled linearly, the resulting *a* parameter (Figure 2) can reflect the effect of the treatment (*T*) on unobtainable mediator values. This is analogous to modeling a binary outcome variable using a linear model specification. To address this, a nonlinear function with lower and upper asymptotes should be applied to the treatment's effect on the mediator. This requires a generalization of the nonlinear model used to analyze dichotomous mediating variables.

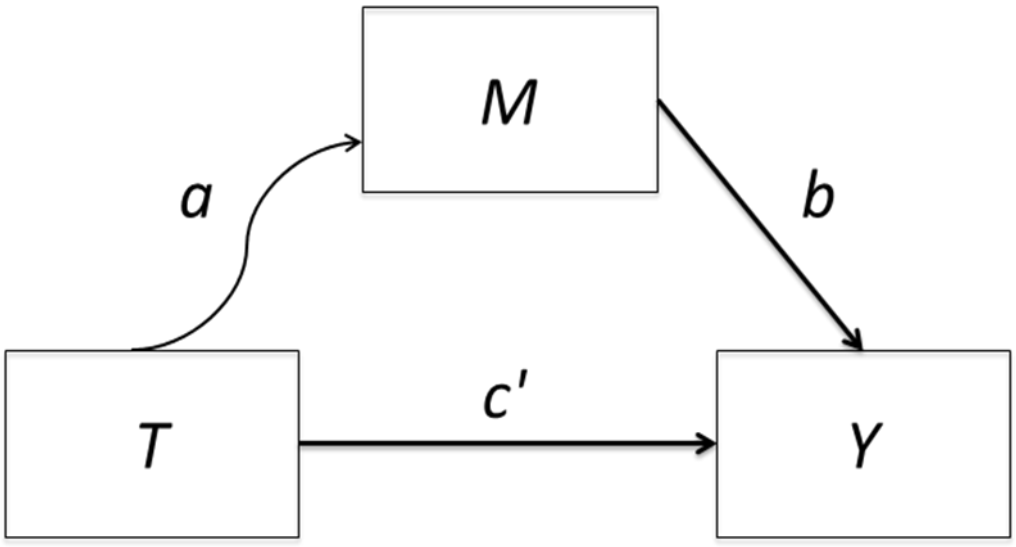
Mediators with floor and ceiling effects require a link function that contains lower and upper asymptotes corresponding to the mediating variables' minimum and maximum

values. A generalized logistic trajectory can be used to model the effect of T on the mediator, M, as follows:

$$M_i = \alpha_1 + \frac{(\alpha_2 - \alpha_1)}{1 + \gamma_0 \exp[-(\gamma_1 T_i)]} + r_i, \tag{50}$$

where M_i represents the value on the mediator for subject *i,* α_1 and α_2 respectively represent lower and upper mediator asymptotes, γ_0 is a pseudo-intercept parameter that influences the point at which the function crosses the ordinate axis, γ_1 is a pseudo-slope parameter that is related to the rate of change of the nonlinear function between its asymptotes, and r_i represents the residual for subject *i*. In contrast to instances involving dichotomous mediators, mediating variables containing upper and lower asymptotes are both continuous and observed. As such, the residual, r_i, is included in the model specification and its variance is directly estimable from the observed data. This precludes the need to utilize the threshold latent variable model discussed above. Figure 4 provides a path analysis diagram of the proposed nonlinear relationship between the treatment and the mediator.

Figure 4: A mediation model in which the relationship between the treatment and the mediator is modeled nonlinearly.



Borrowing an exposition from Singer and Willett (2003), Figure 3 depicts graphs of Equation 50 for various values of γ_0 and γ_1 with α_1 and α_2 set at zero and 100, respectively. Although the γ_0 and γ_1 parameters do not have the same interpretations as in the linear model (see Equations 1 through 3), *M*-axis intercept values are clearly related to γ_0, and the rate at which the function reaches its upper asymptote is clearly related to γ_1. Thus, in keeping with the terminology introduced in Singer and Willett, γ_0 and γ_1 will be referred to as the pseudo-intercept and pseudo-slope parameters, respectively. As can be seen in the graphs in Figure 3, the larger the value of γ_1, the steeper the acceleration of the curve. The larger the value of γ_0, the larger the value of the

M-axis intercept. This function simplifies to the standard logistic regression equation for binary outcomes given $\alpha_1 = 0$, $\alpha_2 = 1$, and $\gamma_0 = 1$.

In practice, the γ_1 parameter must be greater than or equal to zero with a small population variance. In cases where this value is less than zero, the portion of the nonlinear trajectory between the asymptotes requires negative values for the treatment variable. While this is mathematically possible, negative treatment values for continuous variables seem highly unlikely in practice. In the case where this parameter equals zero, Equation 50 simplifies to a constant, resulting in a horizontal line that fails to capture the dynamic relationship between the treatment and the mediator. Additionally, small changes in this parameter will result in noticeably different trajectories. As evidenced in Figure 3, a 0.2 change in the value of the γ_1 parameter results in markedly different trajectories across the selected values of γ_0. This suggests that large variances (particularly in the context of simulation studies), may pose problems for model convergence.

The parameterization of the nonlinear model in Equation 50 is designed for use with single-level datasets. However, this parameterization is easily extended to multilevel mediation analyses. For example, in $1 \rightarrow 1 \rightarrow 1$ mediation, the treatment's effect on the mediator may vary across level-2 units. At level-1, this could be parameterized as follows:

$$M_{ij} = \alpha_1 + \frac{(\alpha_2 - \alpha_1)}{1 + \beta_{(M.T)0j} \exp[-(\beta_{(M.T)1j} T_{ij})]} + r_{ij}, \tag{51}$$

and at level-2:

$$\begin{cases} \beta_{(M.T)0j} = \gamma_{00} + u_{0j} \\ \beta_{(M.T)1j} = \gamma_{10} + u_{1j} \end{cases}, \tag{52}$$

resulting in the following combined equation:

$$M_{ij} = \alpha_1 + \frac{(\alpha_2 - \alpha_1)}{1 + (\gamma_{00} + u_{0j}) + \exp[-(\gamma_{10} + u_{1j})T_{ij}]} + r_{ij}. \tag{53}$$

Here, γ_{00} and γ_{10}, respectively, represent the multilevel extensions of the parameters defined in Equation 50, and u_{0j} and u_{1j} represent their respective random effects. In this design, the model for the outcome could be modeled linearly as in Equations 22 and 23 since a nonlinear *T-M* relationship does not preclude the possibility of a linear *M-Y* relationship. Although the distribution of the mediator will necessarily be truncated at the lower- and upper- asymptotes (α_1 and α_2, respectively), the relationship between the mediator and the outcome may still be linear in nature. While other, nonlinear *M-Y* relationships may be specified, the current study will examine a linear relationship as a initial attempt at estimating the model of interest. The level-1 and -two residuals for Equations 22, 23, and 53 are assumed to be distributed in the same manner as those described for the $1 \rightarrow 1 \rightarrow 1$ mediation models in Equations 20 through 23.

Alternatively, Equation 50 could be extended for use with $2 \rightarrow 1 \rightarrow 1$ mediation analyses according to the following specification at level-1:

$$M_{ij} = \alpha_1 + \frac{(\alpha_2 - \alpha_1)}{1 + \beta_{(M.T)0j} \exp[-(\beta_{(M.T)1j})]} + r_{ij}, \tag{54}$$

and at level-2:

$$\begin{cases} \beta_{(M.T)0j} = \gamma_{00} + u_{0j} \\ \beta_{(M.T)1j} = \gamma_{11} T_j + u_{1j} \end{cases}, \tag{55}$$

resulting in the following combined model:

$$M_{ij} = \alpha_1 \frac{(\alpha_2 - \alpha_1)}{1 + (\gamma_{00} + u_{0j})\{\exp[-(\gamma_{11} T_j + u_{1j})]\}} + r_{ij}. \tag{56}$$

Here, γ_{00} is the pseudo-intercept, γ_{11} is the pseudo-slope, and u_{0j} and u_{1j} respectively represent intercept and slope random effects. Level-1 and -two residuals are distributed as those described for $2 \rightarrow 1 \rightarrow 1$ mediation analyses in Equations 12 and 13. Though the asymptotes α_1 and α_2 may vary across level-2 units, Equations 53 and 56 treat them as fixed effects. To obtain paths *b* and *c* in Figure 2, the model for the outcome could be modeled linearly as in Equations 14 and 15.

This specification may be further extended to include $2 \rightarrow 2 \rightarrow 1$ mediation designs. Here, the treatment and the mediator are measured at level-2, while the outcome is measured at the level of the individual subject (level-1). Since the treatment and the mediator are both measured at the highest level in the data hierarchy, there is no clustering to model for the effect of the treatment on the mediator. Hence, the model for the mediator is:

$$M_j = \alpha_1 + \frac{(\alpha_2 - \alpha_1)}{1 + \gamma_0 \exp[-(\gamma_1 T_j)]} + r_{ij} \tag{57}$$

where γ_0 and γ_1 represent the pseudo-intercept and the pseudo-slope, respectively, and all other parameters are defined as above. The model for the outcome, *Y*, must address the nesting of individuals within level-2 units by utilizing multilevel modeling techniques,

and could be modeled linearly as in Equations 16 and 17. Again, level-1 and -two residuals are distributed as those described for $2 \rightarrow 2 \rightarrow 1$ mediation in Equations 16 and 17.

This series of multilevel logistic models can be used for calculating the indirect effect in one of three ways. First, the *ab* product may be used as an estimate of the indirect effect and tested for statistical significance. However, this conceptualization assumes that the treatment's effect on the mediator is constant across all dosages of the treatment. Furthermore, this conceptualization does not correspond to the mathematical definition of the instantaneous indirect effect (Equation 4). As such, as with binary mediators and outcomes, the *ab* estimate of the indirect effect tends to overestimate the true magnitude of the mediated effect (Li et al., 2007). More generally, the instantaneous indirect effect may be overestimated or underestimated depending on the nonlinear function specified and the location along the nonlinear trajectory where the indirect effect is estimated. Second, the $(c-c')$ estimate of the indirect effect may be utilized, although this estimate is sensitive to the distribution of the treatment dosage. As such, the $(c-c')$ estimate should be used cautiously, if at all (Li et al., 2007). Finally, combining the mathematical definition of the instantaneous indirect effect (Equation 4) with the generalized equation for an asymptotic *M* (Equation 50) yields a third method of estimating the indirect effect. In this case, the instantaneous indirect effect provides an estimate of the indirect effect that depends on the dosage of the treatment variable. This conceptualization allows for a non-constant relationship between *T* and *M* corresponding to the nonlinear trajectory specified in Equation 50. Because the value of the

instantaneous indirect effect depends on the level of the treatment, the instantaneous indirect effect is a form of moderated mediation. As such, it is recommended that the instantaneous indirect effect be calculated at the sample mean of *T* as well as at plus and minus one standard deviation from the mean of *T* to investigate the instantaneous indirect effect at low, moderate, and high dosages of the treatment variable (Hayes & Preacher, 2010). As previously mentioned, the current analysis focuses solely on models containing continuous treatment variables, with the understanding that formulations for binary treatment indicators must be derived separately (Li et al., 2007). Although the use of the instantaneous indirect effect applies to $1 \rightarrow 1 \rightarrow 1$, $2 \rightarrow 1 \rightarrow 1$, and $2 \rightarrow 2 \rightarrow 1$ mediation models, the current study and the following derivation for the instantaneous indirect effect focus only on $1 \rightarrow 1 \rightarrow 1$ designs.

In a $1 \rightarrow 1 \rightarrow 1$ nonlinear multilevel mediation model, as defined in Equations 53, 22, and 23, the instantaneous indirect effect would be:

$$\theta = \left(\frac{\partial M}{\partial T}\right)\left(\frac{\partial Y}{\partial M}\right) = \left\{\frac{\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right\}b. \quad (58)$$

Here, only fixed effects' terms are considered in the derivation of the instantaneous indirect effect. Using this conceptualization, Equation 58 shows that the indirect effect varies as a function of the value on the treatment variable, *T*. With nonlinear mediation as parameterized in Equation 53, the treatment eventually reaches a point of diminishing returns as the function representing the treatment's effect on the mediator approaches its upper asymptote. This point, defined, here, as the level of treatment dosage optimization, corresponds to the point of inflection on the logistic trajectory specified in Equation 53.

The point of treatment dosage optimization holds important implications for optimizing the treatment effect through the mediator when the direct effect of the treatment on the outcome (*c'*) is close to zero. In these cases, the treatment's effect on the outcome is largely a result of the treatment's indirect effect on the outcome through the mediator. The point of treatment dosage optimization thus identifies a level of treatment beyond which the treatment's effect on the mediator (and, consequently, its effect on the outcome) is subject to diminishing returns. Optimization of treatment dosage can be defined mathematically as the value of *T*, at which the rate of change of θ is zero, representing the point of inflection on the logistic trajectory. Mathematically, setting the derivative of θ with respect to *T* equal to zero identifies the point of inflection on the logistic trajectory. First, the derivative of θ is derived as follows:

$$\frac{d\theta}{dT} = [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(\frac{2\gamma_{00}\gamma_{10}\exp[-(2\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3} - \frac{\gamma_{10}\exp[-(\gamma_{10}T)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right). \quad (59)$$

Next, setting $\frac{d\theta}{dT}$ equal to zero and solving for *T* provides the point of treatment dosage optimization, T_O:

$$T_O = \frac{\ln(\gamma_{00})}{\gamma_{10}} \quad (60)$$

Equation 60 quantifies the point of inflection such that the effect of the level of *T* on *M* is at a maximum at this point and starts decreasing for higher values on *T*. Although other considerations, such as the time elapsed between the measurement of *T*, *M*, and *Y*, may influence this value, Equation 60 provides a means of quantifying the point at which

treatment efficacy may begin to level off. See the Appendix for derivations of Equations 58 and 59.

PRODCLIN may be used to establish confidence intervals around the instantaneous indirect effect in Equation 58. To do so, both the value of the first term in Equation 58 as calculated at a specific treatment value and its standard error must be supplied to PRODCLIN. To estimate the standard error of a function, Rao (1973) and Sobel (1982; 1986) present the multivariate delta method for deriving the asymptotic variance of a function of more than one random variable. This method requires pre- and post-multiplying the covariance matrix of the function's parameters by a vector of first partial derivatives of the function with respect to its constituent random variables. This results in a heteroscedastic standard error that depends on a particular treatment dosage. The Appendix provides this derivation, using a first-order Taylor series expansion, for the first term of the instantaneous indirect effect in Equation 58. Using the estimated coefficients, the covariance matrix between them, and a pre-determined treatment value of interest, the value of the first term in Equation 58 and its standard error can then be supplied to PRODCLIN, along with the estimated value of *b* and its standard error. PRODCLIN then provides an asymmetric confidence interval around the instantaneous indirect effect at the specified treatment value.

STATEMENT OF PURPOSE

Mediation analyses in the context of clustered data have typically involved assumption of linear relationships between the constituent mediation model variables. In the presence of mediators containing floor- and/or ceiling-effects, linear models fail to

capture the underlying nature of the data. In such cases, specification of a nonlinear model for the relationship between a treatment and a proposed mediator contains three primary advantages over conventional linear methods. First, mediating constructs often contain absolute maximum and minimum values that are ignored in linear parameterizations. Use of the nonlinear specification suggested here alleviates this concern. Second, utilization of the instantaneous indirect effect as defined by Stolzenberg (1980; Equation 4) allows the assessment of the indirect effect as a function of the dosage of the treatment variable. Indirect effects that vary across the spectrum of treatment dosages allow investigators to specify the treatment's relationship with the mediator at all dosages of the treatment. This form of analysis is unavailable via linear mediation methods as the instantaneous indirect effect is assumed constant across all levels of the treatment. As such, nonlinear methods (when appropriate) provide additional information through their interpretation as moderated mediating relationships. Furthermore, this form of moderated mediation avoids issues with *ab* and $(c - c')$ estimates of the indirect effect commonly encountered when specifying nonlinear relationships. Finally, in instances of partial or complete mediation, nonlinear specification allows for investigation of an optimal amount of treatment dosage. This is particularly relevant to behavioral interventions designed to be administered in organizational settings. This ensures that organizations maximize the effectiveness of their interventions, thus conserving the organization's resources while ensuring that the intervention's participants are not subject to unnecessary and unbeneficial amounts of the intervention.

Although the nonlinear multilevel mediation model described in Equations 53 and 56 provides a mathematical model for a treatment-mediator relationship when the mediator has an upper- and lower-asymptote, the statistical properties and performance of the proposed model are yet to be empirically investigated. To examine use and estimation of the proposed nonlinear multilevel mediation model, a simulation study is proposed to assess the estimation of the instantaneous indirect effect in the presence of a truly nonlinear treatment-mediator relationship. The proposed simulation investigates only the $1 \rightarrow 1 \rightarrow 1$ design described in Equation 53, focusing on conditions common to the behavioral and social sciences as measures of mediating constructs with upper- and lower-asymptotes are commonly encountered in these contexts. The purpose of this study is to compare both linear and nonlinear specifications for the instantaneous indirect effect in the presence of a truly nonlinear treatment mediator relationship and to examine the utility of the empirical-M based PRODCLIN method for establishing confidence intervals around the instantaneous indirect effect. Manipulated conditions will include the nonlinear pseudo-slope parameter, the *b* mediation path, level-1 and level-2 sample sizes, and the intraclass correlation coefficient value.

Chapter 3: Methods

GENERATING MODELS

To examine the estimation of the proposed nonlinear multilevel mediation model, a simulation study was conducted for the $1 \rightarrow 1 \rightarrow 1$ multisite experimental design in which participants are randomly assigned to a treatment dosage and are nested within data collection sites. The SAS programming environment (version 9.2) was used to create simulated datasets fitting two generating models: one for the mediator, *M*, and one for the outcome, *Y*. First, the following generating model was used for the mediator at level-1:

$$M_{ij} = \frac{100}{1 + \beta_{(M.T)0j} \exp[-(\beta_{(M.T)1j} T_{ij})]} + r_{ij}, \tag{61}$$

and at level-2:

$$\begin{cases} \beta_{(M.T)0j} = \gamma_{00} \\ \beta_{(M.T)1j} = \gamma_{10} + u_{1j} \end{cases}, \tag{62}$$

Here, α_1 and α_2 were explicitly fixed at zero and 100, respectively, thereby defining a mediator that takes on values between zero and 100 along with a nonlinear relationship with the treatment dosage variable, *T*. Values for the treatment variable were randomly selected from a uniform distribution with a minimum of zero and a maximum of 50. The value of the pseudo-intercept parameter (γ_{00}) was fixed at 150 (to match the pattern depicted in the relevant γ_{00} = 150 graph in Figure 3). Values for the pseudo-slope parameter (γ_{10}) were specified as either 0.14 or 0.39 for use in calculating the indirect effect. These values represent a subset of the generating values used in Li et al.'s (2007)

examination of the mediating effect with binary mediators. The level-1 random effects (r_{ij}) were sampled from a random normal distribution with a mean of zero and standard deviation equal to 10, approximating the results from Singer and Willett's (2003) applied example regarding the logistic change trajectory. To facilitate convergence, the pseudo-intercept parameter (γ_{00}) was fixed across level-2 units. Additionally, the lack of a residual for this term is meant to simulate the scenario in which participants in a particular intervention must pass a set of global diagnostic or screening criteria, thereby reducing their initial variability, i.e., the variability in the pseudo-slope parameter across level-2 units. The u_{1j} random effect was sampled from a normal distribution with a mean of zero and a standard deviation of 0.2. The level-1 residual was not allowed to covary with the level-2 residual.

The outcome variable (*Y*) was generated according to the following linear multilevel model at level-1:

$$Y_{ij} = \beta_{(Y.MT)0j} + \beta_{1j} M_{ij} + \beta_{2j} T_{ij} + r_{ij}, \tag{63}$$

and at level-2:

$$\begin{cases} \beta_{(Y.MT)0j} = \gamma_{00} + u_{0j} \\ \beta_{1j} = \gamma_{10} \\ \beta_{2j} = \gamma_{20} \end{cases}, \tag{64}$$

where M_{ij} comes from Equations 61 and 62. Here, γ_{10} and γ_{20} represent the *b* and *c'* parameters, respectively. Estimation of this random-intercept model with fixed treatment and mediating effects (Equations 63 and 64) was used to facilitate comparisons with

similarly parameterized models from previous research regarding multisite mediation models (Pituch et al., 2005; Krull & MacKinnon, 1999, 2001; MacKinnon, Warsi, & Dwyer, 1995).

The *b* parameter for use in calculating the indirect effect (γ_{10} in Equation 64) was specified to be zero or 0.39, representing differing degrees for the relationship between the mediator and outcome variable. The intercept parameter in Equation 64 (γ_{00}) was specified to be equal to zero, and the direct effect of the treatment on the outcome accounting for the mediator (*c'*; γ_{20} in Equation 64) was held constant at 0.20. These parameter values stem directly from previous research on nonlinear treatment-mediator relationships (e.g., Li et al., 2007). The level-1 random effect (r_{ij}) was sampled from a normal distribution with a mean of zero and a standard deviation of one. The level-2 random effect (u_{0j}) was sampled from a random normal distribution with a mean of zero and standard deviation calculated in accordance with the ICC associated with the current study condition. The level-1 random effect was not allowed to covary with the level-2 random effect.

Next, the number of sites and the number of participants within sites were varied in a manner similar to those used in Pituch et al. (2005). More specifically, data for 10 and 30 sites were generated and completely crossed with the number of participants in each site being either 15 or 30 for a total of four combinations of sample size conditions. These values are consistent with values observed in applied multisite investigations (e.g., Plewis & Hurry, 1998; Pituch & Miller, 1999).

Finally, values for the residual ICCs were set to either .10 or .20 in the equation for the outcome. The covariance between level-2 residuals was set to zero. These values were chosen to be representative of observed values in educational research.

ESTIMATING MODELS

Once generated, two sets of multilevel models were estimated to assess parameter bias for the estimated mediation model parameters and the instantaneous indirect effect at the mean treatment dosage as well as one standard deviation above and below the mean. These analyses were designed to compare nonlinear and linear model specifications in the presence of a truly nonlinear relationship and to investigate the bias of the resulting parameter and instantaneous indirect effect estimates. The first set of multilevel models specified a nonlinear treatment-mediator relationship and were identical to the model used to generate values on *M* as a function of *T* (see Equations 61 and 62). The model for the outcome was identical to the generating equations for the outcome (Equations 63 and 64). The proposed simulation used SAS PROC NLMIXED to estimate the nonlinear relationship for the mediator, and SAS PROC MIXED was used to estimate the linear model for the outcome. The second set of models specified a linear treatment-mediator relationship according to the following multilevel parameterization at level-1:

$$M_{ij} = \beta_{(M.T)0j} + \beta_{(M.T)1j} T_{ij} + r_{(M.T)ij} \ , \tag{65}$$

And at level-2:

$$\begin{cases} \beta_{(M.T)0j} = \gamma_{00} + u_{0j} \\ \beta_{(M.T)1j} = \gamma_{10} \end{cases} . \tag{66}$$

The model for the outcome was identical for the generating model in Equations 63 and 64. SAS PROC MIXED was used to estimate both linear models.

ANALYSIS

Considering all combinations of conditions, this study entailed a 2x2x2x2x2 fully factorial design, resulting in 32 unique conditions. 1,000 replication datasets were generated for each condition. The instantaneous indirect effect for each replication was calculated using Equation 58 at the mean of *T* and at one standard deviation above and below the mean of *T*. Confidence intervals for the instantaneous indirect effect were calculated using the PRODCLIN program outlined by MacKinnon et al. (2007a) for average treatment dosages as well as at treatment dosages one standard deviation above and below the average. Coverage rates for the proportion of replications containing the true value of the instantaneous indirect effect for each condition were reported, as were the proportion of times the true instantaneous indirect effect lay to the left or to the right of the 95% PRODCLIN confidence intervals.

The relative parameter bias was assessed for the pseudo-slope parameter (γ_{10} in Equation 62) and for the instantaneous indirect effect (Equation 58) at the mean value of *T* and at one standard deviation above and below the mean value of *T*. For conditions containing a true value of zero for the instantaneous indirect effect, absolute bias was reported in place of relative bias. These bias estimates were used to assess estimation of the nonlinear model for the mediator. The relative parameter bias for the *b* parameter (γ_{10} in Equation 64) was compared to previous bias estimates (i.e, Pituch et al., 2006) to

ensure proper model estimation. Relative parameter bias was estimated according to the following equation:

$$B(\hat{\theta})_{\text{Relative}} = \frac{\bar{\hat{\theta}}_i - \theta}{\theta}, \tag{66}$$

where $\bar{\hat{\theta}}_i$ equals the represents the average parameter value across all replications for one study condition (Hoogland & Boomsma, 1998). Hoogland and Boomsma's recommended cutoff for substantial parameter bias ($\left|B(\hat{\theta})\right| \geq 0.05$) was used to assess the severity of the observed bias. Absolute bias was calculated according to the following equation:

$$B(\hat{\theta})_{Absolute} = \bar{\hat{\theta}}_i - \theta \ . \tag{67}$$

The efficiency of the proposed indirect effect estimator was reported as the root mean squared error (RMSE) according to the following equation:

$$RMSE(\theta) = \sqrt{\left(\bar{\hat{\theta}}_i - \theta\right)^2} \ . \tag{68}$$

This value was compared across the linear and nonlinear estimators to compare the efficiency of the proposed estimator across model specifications.

Chapter 4: Results

Results presented below summarize non-convergence rates, parameter bias, root mean squared error, and PRODCLIN confidence interval coverage rates for both the linear and nonlinear estimators across study conditions. Results are compared across estimator type (linear or nonlinear), focusing on differences in the observed ranges for the measure of interest across estimator type. Ranges are also reported across generating study conditions to examine the effects of the manipulated parameters on the reported outcomes.

NON-CONVERGENCE RATES

Table 1 presents non-convergence rates for the three models estimated (linear outcome, linear mediator, and nonlinear mediator) during each replication. These values are not mutually exclusive, as all three models were estimated before the determination was made to delete the replication due to at least one non-convergent model. Subsequent replications were run for each condition to ensure that each condition contained a total of 1,000 replications, which met convergence criteria for all three estimated models. Overall, linear model estimation for both the mediator and the outcome converged more frequently than did the nonlinear model for the mediator. Non-convergence rates for the linear outcome and linear mediator models ranged from zero to 7.9% and from zero to 0.9%, respectively. In general, study conditions with only 10 level-2 units proved more problematic for linear model estimation than did study conditions with 30 level-2 units. Non-convergence rates for the outcome model for conditions containing 10 level-2 units

ranged from zero to 7.9%, while outcome model non-convergence rates for conditions containing 30 level-2 units ranged from zero to 0.3%. Non-convergence rates for the linear mediator model were only problematic for conditions containing a pseudo-slope parameter of 0.39 and 10 level-2 units, with a range of zero to 0.9%. No other conditions encountered convergence issues for the linear mediator model, as convergence rates for all other conditions equaled 100%.

Conditions with an ICC equal to 0.10 were more problematic for the outcome model than were conditions with an ICC equal to 0.20, with non-convergence rates ranging from zero to 7.9% and zero to 2.2%, respectively. The magnitude of the *b* parameter had little effect on both linear outcome and linear mediator model convergence. Non-convergence rates for conditions with a generating value of zero for the *b* parameter ranged from zero to 7.3%. Non-convergence rates for the linear outcome model for conditions with a generating value of 0.39 for the *b* parameter ranged from zero to 7.9%. Non-convergence rates for the outcome model were highest for conditions with low numbers of level-1 (n=15) and level-2 (N=10) units combined with low ICC values (ICC=0.10), with rates ranging from 6.0% to 7.9%.

Non-convergence rates for the nonlinear mediator model were consistently higher than non-convergence rates for either linear model. As displayed in Table 1, nonlinear model non-convergence rates ranged from 9.7% to 44.3%. Conditions with 10 level-2 units converged more frequently than conditions with 30 level-2 units, with non-convergence rates ranging from 11.7% to 26.4% and 9.7% to 44.3%, respectively, with 12 of the 16 30 level-2 unit conditions exceeding the maximum non-convergence rate

observed across all 10 level-2 unit conditions. Non-convergence rate discrepancies are less apparent across level-1 sample sizes, with ranges of 9.7% to 44.3% across conditions with 15 level-1 units and 11.7% to 40.7% across conditions with 30 level-1 units. ICC values had little effect on non-convergence rates; non-convergence rates for conditions with an ICC equal to 0.1 range from 9.7% to 44.3%, while non-convergence rates for conditions with an ICC of 0.2 range from 12.4% to 44.2%.

Non-convergence rates across *b* parameter value conditions were similar, with rates ranging from 11.7% to 44.3% for conditions with a zero generating value for the *b* parameter, and from 9.7% to 42.8% for conditions with a 0.39 generating value for the *b* parameter. Non-convergence rates across pseudo-slope generating conditions were noticeably different, ranging from 9.7% to 32.4% across conditions with a generating pseudo-slope value of 0.14, and from 20.4% to 44.3% across conditions with a generating pseudo-slope value of 0.39. Eight of the 16 conditions containing a pseudo-slope parameter equal to 0.39 had non-convergence rates exceeding the maximum observed value across conditions with a generating pseudo-slope value of 0.14. Conditions with a pseudo-intercept value of 0.39 and a level-2 sample size of 30 were particularly problematic, as all non-convergence rates for these study conditions exceeded 39%.

Summary of non-convergence patterns

In summary, convergence was less problematic for both linear models than for the nonlinear model. Linear model convergence was particularly affected by level-2 sample size and ICC values, with high level-2 sample size conditions and high ICC value conditions converging more frequently. Nonlinear model convergence was most affected

by the number of level-2 units and the generating value of the pseudo-slope parameter. Unlike convergence rate patterns for the linear models, the nonlinear model converged more frequently for low level-2 sample size conditions. The nonlinear model converged more frequently for conditions with a generating value of 0.14 for the pseudo-slope parameter than for conditions with a generating value of 0.39.

Table 1. Non-convergence rates for estimated models.

Condition					Non-convergence Rate		
b	γ_{10}	Clusters	Subjects	ICC	Outcome Model	Linear Mediator Model	Nonlinear Mediator Model
0	0.14	10	15	0.10	0.060	0.000	0.206
				0.20	0.009	0.000	0.176
			30	0.10	0.011	0.000	0.117
				0.20	0.005	0.000	0.197
		30	15	0.10	0.000	0.000	0.311
				0.20	0.000	0.000	0.272
			30	0.10	0.000	0.000	0.324
				0.20	0.000	0.000	0.124
	0.39	10	15	0.10	0.073	0.009	0.264
				0.20	0.022	0.005	0.223
			30	0.10	0.009	0.000	0.214
				0.20	0.002	0.001	0.214
		30	15	0.10	0.003	0.000	0.443
				0.20	0.000	0.000	0.442
			30	0.10	0.000	0.000	0.404
				0.20	0.000	0.000	0.387
0.39	0.14	10	15	0.10	0.079	0.000	0.184
				0.20	0.010	0.000	0.208
			30	0.10	0.011	0.000	0.198
				0.20	0.002	0.000	0.209
		30	15	0.10	0.001	0.000	0.097
				0.20	0.000	0.000	0.124
			30	0.10	0.000	0.000	0.136
				0.20	0.000	0.000	0.307
	0.39	10	15	0.10	0.078	0.003	0.227
				0.20	0.012	0.002	0.220
			30	0.10	0.009	0.000	0.204
				0.20	0.000	0.001	0.230
		30	15	0.10	0.001	0.000	0.416
				0.20	0.000	0.000	0.428
			30	0.10	0.000	0.000	0.407
				0.20	0.000	0.000	0.398

PARAMETER AND RELATIVE PARAMETER BIAS

Table 2 presents relative parameter bias results for the estimation of the pseudo-slope parameter. Overall, relative parameter bias for the pseudo-slope parameter ranged from -1.3% to 6.8%, and was within the recommended 5% cutoff for all but two conditions. Generating values for the *b* parameter and for the ICC values had no effect on pseudo-slope bias, as these values were not used to generate the nonlinear mediator model. Level-2 sample size had a slight impact on pseudo-slope bias, with relative parameter bias value ranging from -1.3% to 6.8% for low level-2 unit conditions, and from -0.5% to 2.0% for high level-2 unit conditions. Level-1 sample size had no noticeable effect on relative parameter bias, ranging from -0.5% to 6.8% for low level-1 sample size conditions, and from -1.3% to 4.1% for high level-1 sample size conditions.

The generating value of the pseudo-slope produced noticeable discrepancies in relative parameter bias values, ranging from -0.5% to 6.8% for conditions with a generating value of 0.14, and from -1.3% to 1.2% for conditions with a generating value of 0.39. Conditions with low level-1 and 2 sample sizes and a generating pseudo-slope value of 0.14 consistently produced the most biased estimates for the pseudo-slope parameter, ranging from 2.5% to 6.8% and exceeding the absolute value of the relative parameter bias for all remaining conditions. Within low level-2 sample size and low pseudo-slope conditions, low level-1 sample size conditions performed worse than did high level-1 sample size conditions, ranging from 4.5% to 6.8% versus from 2.5% to 4.1%, respectively. Two of the four conditions with low pseudo-slope, low level-2, and

low level-1 generating values returned relative parameter bias values in excess of the recommended 0.05 cutoff, with values of 6.8% and 5.2%.

Table 2. Relative parameter bias for the pseudo-slope parameter.

Conditions					
b	γ_{10}	Clusters	Subjects	ICC	Bias
0	0.14	10	15	0.10	**0.068**
				0.20	0.045
			30	0.10	0.032
				0.20	0.042
			15	0.10	0.015
				0.20	0.020
			30	0.10	0.001
				0.20	0.020
	0.39	10	15	0.10	0.012
				0.20	0.012
			30	0.10	0.004
				0.20	0.008
		30	15	0.10	0.001
				0.20	0.006
			30	0.10	0.004
				0.20	0.006
0.39	0.14	10	15	0.10	0.047
				0.20	**0.052**
			30	0.10	0.025
				0.20	0.029
		30	15	0.10	-0.006
				0.20	0.008
			30	0.10	0.012
				0.20	-0.004
	0.39	10	15	0.10	0.005
				0.20	0.005
			30	0.10	-0.013
				0.20	-0.001
		30	15	0.10	0.005
				0.20	0.006
			30	0.10	0.005
				0.20	0.005

**Note*: Bold values indicate relative parameter bias in excess of the recommended 0.05 cutoff.

Tables 3 and 4 present bias results for both linear and nonlinear estimates of the instantaneous indirect effect at the mean value of the treatment and at values one standard deviation above and below the mean treatment value. For conditions with *b* values of zero (and, therefore, true instantaneous indirect effect values of zero), Table 3 presents parameter bias calculations in place of relative parameter bias. Relative parameter bias is reported for all other conditions in Table 4. Given the mathematical definition of the instantaneous indirect effect, the linear estimator provides only one estimate for this value. As such, this value was compared to true values of the instantaneous indirect effect at low, mean, and high treatment values across study conditions. The nonlinear model provides an instantaneous indirect effect that is dependent on the value of the treatment, and therefore provides a unique value for parameter bias evaluation at each of three chosen treatment levels.

Across all conditions, estimation of the *b* parameter was unbiased, suggesting proper linear outcome model estimation across all generating conditions. For conditions with a generating value of zero for the *b* parameter, all parameter bias values were equal to zero to two decimal places. Given this, subsequent results will only be presented and compared for conditions containing a generating value of 0.39 for the *b* parameter (i.e., results from Table 3).

Relative parameter bias for conditions with *b* equal to 0.39

Although few conditions returned relative parameter bias values within Hoogland and Boomsma's (1998) recommended 0.05 cutoff, the nonlinear estimator outperformed the linear estimator for all conditions evaluated at low treatment values. For low

treatment values, relative parameter bias ranged from -7.0% to 68.0% for the nonlinear estimator and from -76.0% to 252% for the linear estimator, suggesting that the linear estimator does a poor job of recovering the true parameter in presence of a truly nonlinear relationship.

Across cluster conditions for the nonlinear estimator, relative parameter bias estimates ranged from -7.0% to 68% for conditions with 10 level-2 units, and from -1.0% to 21.0% for conditions with 30 level-2 units. For the linear estimator at low treatment values, relative parameter bias estimates ranged from -76.0% to 252.0% for conditions containing 10 level-2 units, and from -76.0% to 247.0% for conditions containing 30 level-2 units. The number of level-1 units had little effect on relative parameter bias within models. For the nonlinear model, bias estimates ranged from -6.0% to 68.0% for conditions with 15 level-1 units, and from -7.0% to 60.0% for conditions with 30 level-1 units. Bias estimates were systematically worse for the linear estimator, ranging from -76.0% to 252.0% for conditions with 15 level-1 units and from -76.0% to 249% for conditions with 30 level-1 units.

ICC values also had little effect on the observed relative parameter bias, ranging from -7.0% to 68.0% for the nonlinear estimator in conditions with an ICC equal to 0.10 and from -6.0% to 68.0% for the nonlinear estimator in conditions with an ICC equal to 0.20. For the linear estimator, the relative parameter bias was identical across ICC conditions, ranging from -76.0% to 252.0%. Generating values for the pseudo-slope parameter had the most noticeable effect on relative parameter bias across model type. For the nonlinear estimator, relative parameter bias values ranged from 18.0% to 68% for

pseudo-slope values of 0.14, and from -7.0% to 1.0% for pseudo-slope values of 0.39. The relative parameter bias of the linear estimator ranged from 244.0% to 252.0% for pseudo-slope values of 0.14, and was equal to -76.0% for all conditions with a pseudo-slope value of 0.39.

Table 3. Parameter bias of the instantaneous indirect effect for conditions with a value of zero for the *b* parameter.

Conditions					Low Treatment		Mean Treatment		High Treatment	
b	γ_{10}	Clusters	Subjects	ICC	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
0	0.14	10	15	0.10	-0.0001	-0.0002	-0.0004	-0.0002	-0.0003	-0.0002
				0.20	0.0001	0.0001	0.0001	0.0001	0.0000	0.0001
			30	0.10	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000
				0.20	0.0000	-0.0001	-0.0001	-0.0001	0.0000	-0.0001
		30	15	0.10	0.0000	0.0000	-0.0001	0.0000	0.0000	0.0000
				0.20	0.0001	0.0001	0.0003	0.0001	0.0001	0.0001
			30	0.10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
				0.20	0.0000	0.0000	-0.0001	0.0000	-0.0001	0.0000
	0.39	10	15	0.10	0.0002	0.0001	0.0001	0.0001	0.0000	0.0001
				0.20	-0.0007	-0.0002	-0.0001	-0.0002	0.0000	-0.0002
			30	0.10	0.0000	-0.0001	-0.0001	-0.0001	0.0000	-0.0001
				0.20	-0.0011	-0.0002	0.0000	-0.0002	0.0000	-0.0002
		30	15	0.10	-0.0005	-0.0001	0.0000	-0.0001	0.0000	-0.0001
				0.20	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
			30	0.10	-0.0004	-0.0001	-0.0001	-0.0001	0.0000	-0.0001
				0.20	-0.0002	-0.0001	0.0000	-0.0001	0.0000	-0.0001

Note: "Nonlinear" refers to instantaneous indirect effect estimates using the nonlinear model specification. "Linear" refers to instantaneous indirect effect estimates using the linear model.

The relative parameter bias of the instantaneous indirect effect estimates at mean treatment values exceeded the 0.05 recommended cutoff for both the linear and nonlinear estimators across all conditions. Overall, relative parameter bias of the nonlinear model

ranged from 10.0% to 132.0%, and from -36.0% to 477.0% for the linear model. As at low treatment values, sample size at level-2 mitigated relative parameter bias for the nonlinear estimator, ranging from 15.0% to 132.0% for conditions with 10 level-2 units, and from 10.0% to 36.0% for conditions with 30 level-2 units. The linear estimator was not as noticeably affected by sample size at level-2; bias estimates range from -35.0% to 477.0% for the low sample size condition, and from -36.0% to 472.0% for the high sample size condition. Differences were negligible within model type across level-1 sample size conditions. For the nonlinear model and level-1 sample size of 15, relative parameter bias ranged from 10.0% to 122.0%. In conditions with a level-1 sample size of 30, relative parameter bias for the nonlinear model ranged from 12.0% to 132.0%. For the linear model, relative parameter bias associated with conditions containing 15 level-1 units ranged from -36.0% to 477.0%, and from -36.0% to 473.0% for conditions containing 30 level-1 units.

ICC values had little effect on relative parameter bias for either model type. For the nonlinear model, relative parameter bias for conditions with an ICC equal to 0.10 ranged from 10.0% to 132.0%. For conditions with an ICC value of 0.20, relative parameter bias ranged from 12.0% to 130.0% for the nonlinear estimator. The minimum relative parameter bias for the linear estimator was equal to -36.0% for both ICC conditions, with maximum values of 477.0% and 472.0% for ICC values of 0.10 and 0.20, respectively. Generating values for the pseudo-slope parameter had a noticeable effect on relative parameter bias for both the nonlinear and linear models. For the

nonlinear model, relative parameter bias ranged from 10.0% to 18.0% for conditions with a pseudo-slope equal to 0.14.

Table 4. Relative parameter bias of the instantaneous indirect effect for conditions with a value of 0.39 for the *b* parameter.

Conditions					Low Treatment		Mean Treatment		High Treatment	
b	γ_{10}	Clusters	Subjects	ICC	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
0.39	0.14	10	15	0.10	**0.68**	**2.52**	**0.17**	**-0.34**	**-0.48**	**-0.58**
				0.20	**0.68**	**2.52**	**0.17**	**-0.34**	**-0.47**	**-0.59**
			30	0.10	**0.58**	**2.49**	**0.15**	**-0.35**	**-0.49**	**-0.59**
				0.20	**0.60**	**2.47**	**0.18**	**-0.35**	**-0.49**	**-0.59**
		30	15	0.10	**0.18**	**2.44**	**0.10**	**-0.36**	**-0.31**	**-0.59**
				0.20	**0.21**	**2.45**	**0.14**	**-0.36**	**-0.32**	**-0.59**
			30	0.10	**0.21**	**2.47**	**0.15**	**-0.35**	**-0.30**	**-0.59**
				0.20	**0.18**	**2.44**	**0.12**	**-0.36**	**-0.31**	**-0.60**
	0.39	10	15	0.10	**-0.06**	**-0.76**	**1.22**	**4.77**	**12.44**	**1577.10**
				0.20	**-0.06**	**-0.76**	**1.16**	**4.72**	**17.50**	**1563.49**
			30	0.10	**-0.07**	**-0.76**	**1.32**	**4.73**	**13.91**	**1566.16**
				0.20	**-0.05**	**-0.76**	**1.30**	**4.67**	**18.72**	**1549.48**
		30	15	0.10	-0.01	**-0.76**	**0.36**	**4.71**	**1.62**	**1562.21**
				0.20	0.00	**-0.76**	**0.34**	**4.70**	**1.47**	**1559.16**
			30	0.10	0.01	**-0.76**	**0.31**	**4.72**	**1.52**	**1562.96**
				0.20	0.00	**-0.76**	**0.33**	**4.72**	**1.61**	**1562.91**

Note: "Nonlinear" refers to instantaneous indirect effect estimates using the nonlinear model specification. "Linear" refers to instantaneous indirect effect estimates using the linear model. Bolded values indicate relative parameter bias in excess of the recommended 0.05 cutoff.

When the pseudo-slope value was equal to 0.39, the relative parameter bias of the nonlinear estimator ranged from 31.0% to 132.0%. For the linear model, relative parameter bias for conditions with a pseudo-slope of 0.14 ranged from -36.0% to -34.0%, and from 467.0% to 477.0% for conditions with the pseudo-slope equal to 0.39. This

pattern of relative parameter bias pertaining to the pseudo-slope values for the nonlinear model was the opposite of what was observed for low treatment values.

For high treatment values, relative parameter bias for the nonlinear model was less extreme than the relative parameter bias for the linear model estimates across all study conditions. However, no conditions met the recommended 0.05 cutoff as indicative of acceptable parameter bias. Overall, the relative parameter bias of the nonlinear model ranged from -32.0% to 1,872.0%, and from -60.0% to 157,710.0% for the linear model. As seen with low and mean treatment values, level-2 sample size noticeably affected the relative parameter bias for the nonlinear model. For conditions containing 10 level-2 units, relative parameter bias for the nonlinear model ranged from -49.0% to 1,872.0%. When the generating parameters included 30 level-2 units, the relative parameter bias of the nonlinear model shrank, with a range of -32.0% to 162.0%. The number of level-2 units had little impact on the relative parameter bias of the linear model, with a range of -59.0% to 157,710.0% for conditions with 10 level-2 units, and -60.0% to 156,296.0% for conditions with 30 level-2 units.

Unlike level-2 sample size, the number of level-1 units had little effect on relative parameter bias within model type. For the nonlinear model, conditions with 15 level-1 units returned relative parameter bias values ranging from –48.0% to 1,750.0%, while conditions with 30 level-1 units had relative parameter bias values ranging from -49.0% to 18.72%. For the linear model, relative parameter bias ranged from -59.0% to 157,710.0% for conditions with 15 level-1 units, and from -60.0% to 156,616.0% for conditions with 30 level-1 units.

ICC values also proved to have little effect on relative parameter bias within model type. The relative parameter bias for nonlinear estimator ranged from -49.0% to 1,391.0% and -49.0% to 1,872.0% for low and high ICC values, respectively. Relative parameter bias for the linear estimator ranged from -59.0% to 157,710.0% for conditions with an ICC equal to 0.10, and from -60.0% to 156,349.0% for conditions with an ICC value equal to 0.20. Generating values for the pseudo-slope parameter had a noticeable effect on within model relative parameter bias. For the nonlinear model, relative parameter bias ranged from -49.0% to -30.0% for conditions with a pseudo-slope value of 0.14, and from 147.0% to 1,872.0% for conditions with a pseudo-slope value of 0.39. For the linear model, relative parameter bias values ranged from -60.0% to -58.0% for a generating pseudo-slope value of 0.14, and from 154,948.0% to 157,710.0% for conditions with a pseudo-slope value of 0.39. The pattern of relative parameter bias values for the nonlinear model across pseudo-slope conditions mirrored what was observed for low treatment values.

Summary of bias patterns

In summary, conditions with a *b* parameter of zero produced parameter bias values equal to zero (to two decimal places) for both the linear and nonlinear estimator across all generating values. For conditions with a non-zero *b* parameter, the nonlinear estimator outperformed the linear estimator across all study conditions. The nonlinear estimator was most affected by pseudo-slope values and sample size at level-2, although only low treatment conditions with a pseudo-slope value of 0.39 and a *b* parameter value of 0.39 returned relative parameter bias values within or near Hoogland and Boomsma's

(1998) recommended threshold. For mean and high treatment values, the relative parameter bias for the nonlinear model was lowest for conditions containing a pseudo-slope generating value of 0.14. For low treatment values, the relative parameter bias was lowest for conditions containing a pseudo-slope generating value of 0.39. Across the three examined treatment values, conditions with 30 level-2 units returned less biased results within a specific treatment value than did conditions containing 10 level-2 units. Combined with the results from Table 1, this suggests that, although the nonlinear model was more difficult to estimate with a larger number of level-2 units, it was less biased in the presence of more level-2 units when a solution was computationally feasible. The linear estimator was substantially biased across all conditions, never falling below the Hoogland and Boomsma's 0.05 cutoff.

ROOT MEAN SQUARED ERROR (RMSE)

Table 5 shows RMSE values for the instantaneous indirect effect at low, mean, and high treatment values for both the linear and nonlinear model specifications. Patterns of RMSE ranges differed noticeably across the *b* equals zero and *b* equals 0.39 conditions. To facilitate comparisons with the preceding exposition on bias patterns, RMSE results will be presented separately for the two *b* conditions. All values will be reported to two decimal places.

RMSE values for conditions with *b* equal to zero

For conditions with a *b* parameter of zero, the linear estimator proved to be as efficient as or more efficient than the nonlinear estimator when controlling for other

study conditions. These differences, however, were negligible across all comparisons, with a maximum absolute difference of 0.03 between efficiency measures for the linear and nonlinear estimators. For all *b* equals zero conditions, RMSE values for the linear estimator ranged from zero to 0.01, and from zero to 0.03 for the nonlinear estimator. Examining the efficiency of the estimator at low treatment values, level-2 sample size had minimal effect on the efficiency of each estimator. For the nonlinear estimator, RMSE values ranged from zero to 0.03 and from zero to 0.02 for low and high level-2 unit conditions, respectively. For the linear estimator, RMSE values ranged from zero to 0.01 for the 10 level-2 unit condition, and were equal to 0.00 for the linear model across all 30 level-2 unit conditions. These RMSE ranges across low and high level-2 unit conditions were identical to those observed across low and high level-1 unit conditions. ICC values had no effect within model type on estimator efficiency. For the nonlinear model at low treatment values, the RMSE ranged from zero to 0.03 for both ICC value conditions. Similarly, for the linear model at low treatment values, the RMSE ranges from zero to 0.01 across generating ICC value conditions. Pseudo-slope generating values resulted in minor RMSE discrepancies across pseudo-slope generating values. For conditions with a pseudo-slope value of 0.14, both the linear and nonlinear estimators RMSE values were equal to zero across all other study conditions for low treatment values. For conditions with pseudo-slope values of 0.39, the RMSE of the nonlinear estimator ranged from 0.01 to 0.03, and from zero to 0.01 for the linear estimator.

For mean treatment values, the RMSE estimates ranged from zero to 0.01 across conditions for both the linear and nonlinear estimator. Level-2 sample size had no effect

on RMSE ranges for the nonlinear estimator at mean treatment values, ranging from zero to 0.01 for both low and high level-2 sample size conditions. For the linear estimator, RMSE values ranged from zero to 0.01 in the 10 level-2 unit conditions and are equal to zero across the 30 level-2 unit conditions. RMSE ranges across level-1 conditions were identical to those described for level-2 conditions. ICC values had no effect on RMSE ranges for both the linear and nonlinear estimator, ranging from zero to 0.01 for both estimators across both ICC conditions. Across low and high generating values for the pseudo-slope parameter at mean treatment values, the RMSE of the nonlinear estimator ranged from zero to 0.01 for conditions with a generating value of 0.14, and was equal to zero for conditions with a generating value of 0.39. This pattern was reversed for the linear estimator at mean treatment values, with an RMSE of zero for conditions with a pseudo-slope generating value of 0.14 and an RMSE range of zero to 0.01 for conditions with a pseudo-slope generating value of 0.39.

At high treatment values for conditions with a generating *b* parameter value of zero, the RMSE of both the linear and the nonlinear estimators ranged for zero to 0.01. Within level-2 sample size conditions, RMSE ranges were identical for both estimators, ranging from zero to 0.01 in the low level-2 sample size condition and were equal zero in all high level-2 sample size conditions. Considering level-1 sample size conditions at high treatment values, RMSE ranges for both estimators were identical for the low level-1 sample size condition, ranging from zero to 0.01. For the high level-1 sample size condition, the RMSE ranged from zero to 0.01 for the nonlinear estimator and was equal to zero across other study conditions for the linear estimator. Generating ICC values had

no effect on the range of RMSE values across estimator type at high treatment values, ranging from zero to 0.01 for both the linear and nonlinear model across both ICC conditions. Examining pseudo-slope generating parameters at high treatment values, RMSE values for conditions with a pseudo-slope parameter equal to 0.14 for the nonlinear estimator ranged from zero to 0.01 for the nonlinear estimator, and were equal to zero for the linear estimator. For conditions with a generating pseudo-slope parameter of 0.39, RMSE values for the nonlinear estimator were equal to zero across all other study conditions, and range from zero to 0.01 for the linear estimator.

RMSE values for conditions with *b* equal to 0.39

RMSE patterns for conditions with a non-zero instantaneous indirect effect were more complicated than for conditions with a zero instantaneous indirect effect. In general, the nonlinear estimator proved to be more efficient than the linear estimator, with overall ranges of 0.12 to 1.18 and 0.31 to 2.40, respectively. However, the linear estimator did outperform the nonlinear estimator for certain combinations of study conditions at specific treatment values. These conditions will be emphasized as the results are presented below.

At low treatment values, the nonlinear estimator was more efficient than the linear estimator across all study conditions with a generating value of 0.39 for the *b* parameter. RMSE values for the nonlinear estimator ranged from 0.12 to 1.18, and from 0.38 to 2.40 for the linear estimator for low treatment values. Considering the number of generating level-2 units, RMSE values for the nonlinear estimator ranged from 0.28 to 1.18 for conditions containing 10 level-2 units, and from 0.12 to 0.80 for conditions containing 30

level-2 units. This suggests that an increase in level-2 units results in better efficiency for the nonlinear estimator at low treatment values. In contrast, the number of level-2 units had little effect on the efficiency of the linear estimator at low treatment values, with RSME values ranging from 0.40 to 2.40 for conditions with 10 level-2 units, and from 0.38 to 2.39 for conditions with 30 level-2 units.

Table 5. RMSE values for all conditions.

Condition					Low Treatment		Mean Treatment		High Treatment	
b	γ_{10}	Clusters	Subjects	ICC	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
0	0.14	10	15	0.10	0.00	0.00	0.01	0.00	0.01	0.00
				0.20	0.00	0.00	0.01	0.00	0.01	0.00
			30	0.10	0.00	0.00	0.01	0.00	0.00	0.00
				0.20	0.00	0.00	0.01	0.00	0.01	0.00
		30	15	0.10	0.00	0.00	0.00	0.00	0.00	0.00
				0.20	0.00	0.00	0.01	0.00	0.00	0.00
			30	0.10	0.00	0.00	0.00	0.00	0.00	0.00
				0.20	0.00	0.00	0.00	0.00	0.00	0.00
	0.39	10	15	0.10	0.03	0.01	0.00	0.01	0.00	0.01
				0.20	0.03	0.01	0.00	0.01	0.00	0.01
			30	0.10	0.02	0.00	0.00	0.00	0.00	0.00
				0.20	0.02	0.00	0.00	0.00	0.00	0.00
		30	15	0.10	0.01	0.00	0.00	0.00	0.00	0.00
				0.20	0.02	0.00	0.00	0.00	0.00	0.00
			30	0.10	0.01	0.00	0.00	0.00	0.00	0.00
				0.20	0.01	0.00	0.00	0.00	0.00	0.00
0.39	0.14	10	15	0.10	0.31	0.40	0.71	0.31	0.76	0.76
				0.20	0.32	0.40	0.70	0.31	0.75	0.76
			30	0.10	0.28	0.40	0.71	0.31	0.77	0.76
				0.20	0.28	0.40	0.72	0.31	0.77	0.77
		30	15	0.10	0.13	0.38	0.59	0.30	0.54	0.77
				0.20	0.13	0.38	0.60	0.30	0.54	0.76
			30	0.10	0.12	0.38	0.59	0.30	0.52	0.76
				0.20	0.12	0.38	0.59	0.30	0.54	0.77
	0.39	10	15	0.10	1.15	2.39	0.43	0.63	0.03	0.76
				0.20	1.15	2.39	0.41	0.62	0.05	0.75
			30	0.10	1.15	2.39	0.43	0.62	0.03	0.75
				0.20	1.18	2.40	0.44	0.62	0.05	0.75
		30	15	0.10	0.80	2.39	0.17	0.62	0.00	0.75
				0.20	0.80	2.39	0.16	0.62	0.00	0.75
			30	0.10	0.76	2.39	0.17	0.62	0.00	0.75
				0.20	0.76	2.39	0.17	0.62	0.00	0.75

Note: "Nonlinear" refers to instantaneous indirect effect estimates using the nonlinear model specification. "Linear" refers to instantaneous indirect effect estimates using the linear model.

The number of level-1 units had little effect on the efficiency of either estimator at low treatment values. RMSE values for the nonlinear estimator ranged from 0.13 to 1.15 for conditions containing 15 level-1 units and from 0.12 to 1.18 for conditions containing 30 level-1 units. For the linear estimator, RMSE values ranged from 0.38 to 2.39 for low level-1 unit conditions and from 0.38 to 2.40 for high level-1 unit conditions. ICC values also had little effect on the efficiency of either estimator. RMSE values for the nonlinear estimator ranged from 0.12 to 1.15 and from 0.12 to 1.18 for low and high ICC values, respectively. For the linear estimator at low treatment values, RMSE values ranged from 0.38 to 2.39 for low ICC conditions and from 0.38 to 2.40 for high ICC conditions. Pseudo-slope generating values noticeably affected RMSE ranges for both the linear and nonlinear estimators at low treatment values. For the nonlinear estimator, RMSE values ranged from 0.12 to 0.32 for low pseudo-slope values and from 0.76 to 1.18 for high pseudo-slope values. RMSE values for the linear estimator at low treatment values ranged from 0.38 to 0.40 for conditions with a pseudo-slope value of 0.14, and from 2.39 to 2.40 for conditions with a pseudo-slope value of 0.39.

For mean treatment values in conditions with a b parameter equal to 0.39, the linear estimator was more efficient than the nonlinear estimator for approximately half of the remaining study conditions. Overall, RMSE values for the nonlinear estimator ranged from 0.16 to 0.72, and from 0.30 to 0.63 for the linear estimator. Level-2 sample size affected RMSE ranges more noticeably for the nonlinear estimator than for the linear estimator. RMSE values for the nonlinear estimator at mean treatment values ranged from 0.41 to 0.72 for low level-2 sample size conditions, and from 0.16 to 0.60 for high

sample size conditions. RMSE ranges for the linear estimator were similar across level-2 unit conditions, ranging from 0.31 to 0.63 for low level-2 unit conditions and from 0.30 to 0.62 for high level-2 unit conditions.

Level-1 sample size had little effect on the efficiency of either estimator at mean treatment values. For the nonlinear estimator, RMSE values ranged from 0.16 to 0.71 for low level-1 unit conditions and from 0.17 to 0.72 for high level-1 unit conditions. The efficiency of the linear estimator at mean treatment values ranged from 0.30 to 0.63 and from 0.30 to 0.62 for low and high level-1 unit conditions. RMSE ranges across ICC conditions within model type were nearly equivalent at mean treatment values. For the nonlinear estimator, RMSE values ranged from 0.17 to 0.71 for low ICC conditions and from 0.16 to 0.72 for high ICC conditions. RMSE values for the linear estimator ranged from 0.30 to 0.63 for low ICC values and from 0.30 to 0.62 for high ICC values. Pseudo-slope values had the most noticeable effect on efficiency across model type at mean treatment values. For conditions with pseudo-slope values of 0.14 and b values of 0.39, the linear estimator was more efficient than the nonlinear estimator across all other study conditions. Here, the RMSE values of the nonlinear estimator ranged from 0.59 to 0.72, while the RMSE values of the linear estimator ranged from 0.30 to 0.31. This pattern reversed for conditions with pseudo-slope values of 0.39, as the nonlinear estimator was more efficient than the linear estimator across all other study conditions with this generating pseudo-slope value and with true b parameters equal to 0.39. For these conditions, the RMSE values of the pseudo-slope parameter ranged from 0.16 to 0.44 for the nonlinear estimator, and from 0.62 to 0.63 for the linear estimator.

For high treatment values in conditions with a *b* generating value of 0.39, the nonlinear estimator was more efficient than the linear estimator for all but one condition and was only negligibly less efficient in the condition in question. Overall, RMSE values for the nonlinear estimator ranged from zero to 0.77, and from 0.75 to 0.77 for the linear estimator. RMSE values for the nonlinear estimator at high treatment values across Level-2 sample size varied slightly, ranging from 0.03 to 0.77 for low level-2 size conditions and from zero to 0.54 for high level-2 unit conditions. For the linear estimator, RMSE values at high treatment values ranged from 0.75 to 0.77 for both low and high level-2 unit conditions. Considering the effect of level-1 sample size, for the nonlinear model at high treatment values, RMSE values ranged from zero to 0.76 and from zero to 0.77 for low and high level-1 unit conditions, respectively. For the linear estimator at high treatment values, RMSE values ranged from 0.75 to 0.77 for both low and high level-1 unit conditions. Examining generating ICC values, RMSE ranges at high treatment values for the nonlinear model went from zero to 0.77 for both low and high ICC conditions. RMSE values for the linear estimator ranged from 0.75 to 0.77 for both low and ICC conditions. For high treatment values, RMSE ranges for the nonlinear estimator were more affected by the pseudo-slope parameter than were RMSE ranges for the linear estimator. For the nonlinear estimator, RMSE values ranged from 0.52 to 0.77 for the conditions containing a pseudo-slope parameter of 0.14, and from zero to 0.05 for conditions containing a pseudo slope parameter of 0.39. For the linear estimator, RMSE values ranged from 0.76 to 0.77 for low pseudo-slope conditions and from 0.75 to 0.76 for high pseudo-slope conditions.

Across all conditions with a generating value of 0.39 for the *b* parameter for high treatment values, two specific conditions provided equivalent or nearly equivalent RMSE values across model type. With a pseudo-slope of 0.14, level-2 sample size of 10, level-1 sample size of 30, and an ICC of 0.10, the linear estimator was negligibly more efficient than the nonlinear estimator, with RMSE of 0.76 compared to the nonlinear estimators RMSE of 0.77. With a pseudo-slope of 0.14, level-2 sample size of 10, level-1 sample size of 30, and an ICC of 0.20, the linear and nonlinear estimators were equally efficient, both producing an RMSE of 0.77.

Summary of RMSE patterns

In summary, RMSE values differed only negligibly for the linear and nonlinear estimators for conditions with a generating value of zero for the *b* parameter. For conditions with a generating value of 0.39 for the *b* parameter, the nonlinear estimator was more efficient for conditions with high values for the number of level-2 units across treatment values. The number of level-2 generating units had no effect on the efficiency of the linear estimator. For low treatment values, both the linear and nonlinear estimators were most efficient for conditions with a pseudo-slope generating value of 0.14. For mean treatment values, the linear estimator was most efficient for conditions with a pseudo-slope generating values of 0.14, while the nonlinear estimator was most efficient for conditions with a pseudo-slope generating value of 0.39. At high treatment values, conditions with a generating pseudo-slope value of 0.39 produced lower RMSE values than did conditions with a generating value of 0.14 for the pseudo-slope parameter. The value of the pseudo-slope parameter had no effect on the efficiency of the linear

estimator. The linear estimator was slightly more efficient than the linear estimator at high treatment values for the condition with a pseudo-slope value of 0.14, a level-2 sample size of 10, and level-1 sample size of 30, an ICC of 0.10, and a *b* parameter of 0.39. For the condition with a pseudo-slope value of 0.14, a level-2 sample size of 10, a level-1 sample size of 30, an ICC of 0.20, and a *b* parameter of 0.39, the linear and nonlinear estimators were equally efficient at high treatment values. The nonlinear estimator was as efficient as or more efficient than the linear estimator for all other conditions across treatment values.

PRODCLIN COVERAGE RATES

Tables 6, 7, and 8 present PRODCLIN coverage rates for both the linear and nonlinear estimator at low, medium, and high treatment values, respectively. As observed with both parameter bias and RMSE values, PRODCLIN coverage rates differed markedly depending on whether the generating value of the *b* parameter was equal to zero or to 0.39, respectively corresponding to zero and non-zero values for the instantaneous indirect effect at the three predetermined treatment values. Consequently, PRODCLIN results will be presented separately for generating values of the *b* parameter.

Conditions with *b* equal to zero

Tables 6, 7, and 8 present PRODCLIN confidence interval estimates for both the linear and nonlinear estimators at low, mean, and high treatment values, respectively. For conditions with a generating value of zero for the *b* parameter, PRODCLIN coverage rates for the linear estimator ranged from 93% to 96% at all three treatment values. For the nonlinear estimator, coverage rates were equal to 100% for low, mean, and high

treatment values. There was a lack of variability in PRODCLIN coverage rates for both the linear and nonlinear estimators for conditions with a generating value of zero for the *b* parameter.

Conditions with *b* equal to 0.39

For conditions having *b* equal to 0.39 at low treatment values, the nonlinear estimator provided superior coverage to that of the linear estimator for low treatment values. As shown in Table 6, coverage rates for the linear estimator reached a maximum of 1% for exactly one condition (pseudo-slope of 0.14, 10 level-2 units, 15 level-1 units, and an ICC of 0.10); all other conditions produced coverage rates of 0% for the linear estimator. For conditions where the linear confidence interval did not include the true instantaneous indirect effect at low treatment values, the true value of the instantaneous indirect effect lay either to the left or to right for all replications within that condition. In contrast, coverage rates for the nonlinear estimator range from 32% to 87% across conditions for low treatment values.

For the nonlinear estimator at low treatment values, true instantaneous indirect effect values that were not contained in the PRODCLIN confidence interval always fell to the right of estimated confidence interval, doing so between 13% and 68% of the time. Coverage rate ranges for low level-2 sample size conditions were smaller than for high level-2 sample size conditions for the nonlinear estimator, ranging from 56% to 87% for low level-2 sample size conditions, and from 32% to 77% for high sample size conditions. For low level-2 unit conditions, the true value of the instantaneous indirect

effect fell to the right of the PRODCLIN confidence interval between 13% and 44% of the time, and between 23% and 68% of the time for high level-2 unit conditions.

Level-1 sample size had little effect on nonlinear coverage rates, ranging from 32% to 87% for low level-1 sample size conditions, and from 32% to 85% for high level-1 sample size conditions. For replications in which the true value of the instantaneous indirect effect was not covered by the nonlinear estimator, true values fell to the left of the coverage rate between 13% and 68% of the time for low level-1 sample size conditions, and between 15% and 68% of the time for high level-1 sample size conditions.

ICC values had a similarly negligible effect on nonlinear coverage rates at low treatment values, ranging from 32% to 85% for low ICC conditions, and from 32% to 87% for high ICC conditions. For replications in which the true value of the instantaneous indirect effect was not covered by the nonlinear PRODCLIN confidence interval, the true value fell to the left of the confidence interval between 15% and 68% of the time for low ICC conditions, and between 13% and 68% of the time for high ICC conditions.

Pseudo-slope values had the most noticeable effect on both the linear and nonlinear estimators. For conditions with a generating pseudo-slope value of 0.14 at low treatment values, the linear estimator fell to the left of PRODCLIN coverage rate between 99% and 100% of the time (falling within the coverage rate exactly 1% of the time for the afore mentioned specific study condition). In contrast, the nonlinear estimator fell with the PRODCLIN coverage rate between 32% and 59% of the time for pseudo-slope

generating conditions of 0.14, with the true value of the instantaneous indirect effect never falling to the left of the nonlinear PRODCLIN confidence interval. For conditions with a generating pseudo-slope value of 0.39 at low treatment values, true values of the instantaneous indirect effect always fell to the right of the linear PRODCLIN coverage rates. For the nonlinear estimator at low treatment values, PRODCLIN coverage rates range from 72% to 87%, with true values never falling to the left of the estimated confidence interval.

Table 6. Coverage rates for confidence intervals around low treatment values.

Conditions					Left		Coverage		Right	
b	γ_{10}	Clusters	Subjects	ICC	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
0	0.14	10	15	0.10	0.00	0.03	1.00	0.93	0.00	0.04
				0.20	0.00	0.04	1.00	0.93	0.00	0.04
			30	0.10	0.00	0.04	1.00	0.93	0.00	0.03
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
		30	15	0.10	0.00	0.03	1.00	0.95	0.00	0.02
				0.20	0.00	0.02	1.00	0.95	0.00	0.03
			30	0.10	0.00	0.02	1.00	0.94	0.00	0.04
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
	0.39	10	15	0.10	0.00	0.03	1.00	0.95	0.00	0.02
				0.20	0.00	0.02	1.00	0.94	0.00	0.04
			30	0.10	0.00	0.02	1.00	0.94	0.00	0.04
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
		30	15	0.10	0.00	0.02	1.00	0.95	0.00	0.03
				0.20	0.00	0.02	1.00	0.96	0.00	0.02
			30	0.10	0.00	0.03	1.00	0.94	0.00	0.03
				0.20	0.00	0.03	1.00	0.95	0.00	0.03
0.39	0.14	10	15	0.10	0.00	0.99	0.57	0.01	0.43	0.00
				0.20	0.00	1.00	0.59	0.00	0.41	0.00
			30	0.10	0.00	1.00	0.56	0.00	0.44	0.00
				0.20	0.00	1.00	0.58	0.00	0.42	0.00
		30	15	0.10	0.00	1.00	0.32	0.00	0.68	0.00
				0.20	0.00	1.00	0.35	0.00	0.65	0.00
			30	0.10	0.00	1.00	0.36	0.00	0.64	0.00
				0.20	0.00	1.00	0.32	0.00	0.68	0.00
	0.39	10	15	0.10	0.00	0.00	0.85	0.00	0.15	1.00
				0.20	0.00	0.00	0.87	0.00	0.13	1.00
			30	0.10	0.00	0.00	0.85	0.00	0.15	1.00
				0.20	0.00	0.00	0.84	0.00	0.16	1.00
		30	15	0.10	0.00	0.00	0.72	0.00	0.28	1.00
				0.20	0.00	0.00	0.74	0.00	0.26	1.00
			30	0.10	0.00	0.00	0.74	0.00	0.26	1.00
				0.20	0.00	0.00	0.77	0.00	0.23	1.00

Note: "Nonlinear" refers to instantaneous indirect effect estimates using the nonlinear model specification. "Linear" refers to instantaneous indirect effect estimates using the linear model.

PRODCLIN confidence interval coverage rates for mean treatment values are displayed in Table 7. Overall, coverage rates for the nonlinear estimator ranged from 39% to 83%, and from zero to 11% for the linear estimator. When the true value of the instantaneous indirect effect was not contained in the linear PRODCLIN confidence interval, the true value lay only to the right or left of the confidence interval within each study condition. For the nonlinear estimator, in contrast, the true value of the instantaneous indirect effect fell only to the right of the PRODCLIN confidence interval across study conditions for replications in which the nonlinear PRODCLIN confidence interval does not include the true value of the instantaneous indirect effect.

At mean treatment values, coverage rate ranges differed across level-2 sample size conditions for both the linear and nonlinear estimators. For the linear estimator, PRODCLIN coverage rates ranged from 0% to 11% for low level-2 sample size conditions, and from 0% to 1% for high level-2 sample size conditions. For the nonlinear estimator, coverage rates ranged from 57% to 83% for low level-2 unit conditions, and from 39% to 83% for high level-2 unit conditions. PRODCLIN coverage rates were not noticeably different across level-1 sample size conditions. For the linear estimator at mean treatment values, coverage rates ranged from 0% to 11% for low level-1 unit conditions, and from 0% to 6% across high level-1 unit conditions. For the nonlinear estimator at mean treatment values, PRODCLIN coverage rates ranged from 43% to 83% across low level-1 unit conditions, and from 39% to 83% across high level-1 unit conditions.

ICC values also negligibly impacted coverage rates for both the linear and nonlinear estimators. For the linear estimator at low treatment values, coverage rates ranged from 0% to 11% and from 0% to 9% for low and high ICC conditions, respectively. For the nonlinear estimator, coverage rates ranged from 39% to 83% across low ICC conditions, and from 41% to 83% across high ICC conditions. Pseudo-slope generating values provided the most noticeable coverage rate differences for both the linear and nonlinear estimators. For the linear estimator at mean treatment values, coverage rates ranged from 0% to 11% for conditions with a generating pseudo-slope value of 0.14, and were equal to 0% for all conditions with a generating pseudo-slope value of 0.39. Coverage rates for the nonlinear estimator at mean treatment values ranged from 80% to 83% for low pseudo-slope conditions, and from 39% to 60% for high pseudo-slope conditions.

Coverage rates for high treatment values are displayed in Table 8. As was found for estimates at both low and mean treatment values, coverage rates for the nonlinear estimator outperformed coverage rates for the linear estimator for all conditions. For the linear estimator, coverage rates were equal to 0% across all study conditions. Coverage rates ranged from 1% to 71% for the nonlinear estimator. Examining the effect of the number of level-2 units on nonlinear coverage rates at high treatment values, coverage rates ranged from 10% to 71% for low level-2 unit conditions, and from 1% to 67% across high level-2 unit conditions.

The number of level-1 units had little effect on nonlinear coverage rates, ranging from 1% to 71% for low level-1 unit conditions, and from 1% to 70% for high level-1

unit conditions. ICC values had similarly negligible effect, with nonlinear coverage rates ranging from 1% to 70% across low ICC conditions, and from 1% to 71% across high ICC conditions. Pseudo-slope generating values had the most noticeable effect on nonlinear estimator coverage rates, ranging from 61% to 71% for conditions with a pseudo-slope generating value of 0.14, and from 1% to 13% for conditions with a pseudo-slope generating value of 0.39. For replications in which the nonlinear PRODCLIN coverage rate did not include the true value of the instantaneous indirect effect at high treatment values, the true value fell only to the right of the confidence interval for conditions with a pseudo-slope generating value of 0.14. For conditions with a pseudo-slope generating value of 0.39, the true value fell to the left of the nonlinear PRODCLIN confidence interval between 21% and 33% of the time, and to the right of the PRODCLIN confidence interval between 57% and 77% of the time.

Summary of PRODCLIN coverage rates

In summary, for conditions with a generating value of 0.39 for the *b* parameter, coverage rates for the nonlinear estimator always exceeded coverage rates for the linear estimator. At low treatment values, coverage rates were at most 1% for the linear estimator. Coverage rates for the nonlinear estimator for conditions with a 0.39 *b* parameter were larger for conditions with fewer level-2 units across treatment values. Conditions with a lower value for the pseudo-slope parameter produced higher coverage rates for the nonlinear estimator than did conditions with higher pseudo-slope values for mean and high treatment values for conditions with a 0.39 *b* parameter. For low treatment

values, conditions with a higher value for the pseudo-slope parameter produced higher coverage rates than did conditions with lower pseudo-slope values.

Table 7. Coverage rates for confidence intervals around mean treatment values.

Conditions					Left		Coverage		Right	
b	γ_{10}	Clusters	Subjects	ICC	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
0	0.14	10	15	0.10	0.00	0.03	1.00	0.93	0.00	0.04
				0.20	0.00	0.04	1.00	0.93	0.00	0.04
			30	0.10	0.00	0.04	1.00	0.93	0.00	0.03
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
		30	15	0.10	0.00	0.03	1.00	0.95	0.00	0.02
				0.20	0.00	0.02	1.00	0.95	0.00	0.03
			30	0.10	0.00	0.02	1.00	0.94	0.00	0.04
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
	0.39	10	15	0.10	0.00	0.03	1.00	0.95	0.00	0.02
				0.20	0.00	0.02	1.00	0.94	0.00	0.04
			30	0.10	0.00	0.02	1.00	0.94	0.00	0.04
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
		30	15	0.10	0.00	0.02	1.00	0.95	0.00	0.03
				0.20	0.00	0.02	1.00	0.96	0.00	0.02
			30	0.10	0.00	0.03	1.00	0.94	0.00	0.03
				0.20	0.00	0.03	1.00	0.95	0.00	0.03
0.39	0.14	10	15	0.10	0.00	0.00	0.81	0.11	0.19	0.89
				0.20	0.00	0.00	0.83	0.09	0.17	0.90
			30	0.10	0.00	0.00	0.80	0.06	0.20	0.94
				0.20	0.00	0.00	0.80	0.05	0.20	0.95
		30	15	0.10	0.00	0.00	0.81	0.00	0.19	1.00
				0.20	0.00	0.00	0.80	0.01	0.20	1.00
			30	0.10	0.00	0.00	0.83	0.00	0.17	1.00
				0.20	0.00	0.00	0.81	0.00	0.19	1.00
	0.39	10	15	0.10	0.01	1.00	0.57	0.00	0.43	0.00
				0.20	0.00	1.00	0.58	0.00	0.42	0.00
			30	0.10	0.00	1.00	0.60	0.00	0.39	0.00
				0.20	0.01	1.00	0.57	0.00	0.42	0.00
		30	15	0.10	0.00	1.00	0.43	0.00	0.57	0.00
				0.20	0.00	1.00	0.43	0.00	0.57	0.00
			30	0.10	0.00	1.00	0.39	0.00	0.61	0.00
				0.20	0.00	1.00	0.41	0.00	0.59	0.00

Note: "Nonlinear" refers to instantaneous indirect effect estimates using the nonlinear model specification. "Linear" refers to instantaneous indirect effect estimates using the linear model.

Table 8. Coverage rates for confidence intervals around high treatment values.

Conditions					Left		Coverage		Right	
b	γ_{10}	Clusters	Subjects	ICC	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
0	0.14	10	15	0.10	0.00	0.03	1.00	0.93	0.00	0.04
				0.20	0.00	0.04	1.00	0.93	0.00	0.04
			30	0.10	0.00	0.04	1.00	0.93	0.00	0.03
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
		30	15	0.10	0.00	0.03	1.00	0.95	0.00	0.02
				0.20	0.00	0.02	1.00	0.95	0.00	0.03
			30	0.10	0.00	0.02	1.00	0.94	0.00	0.04
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
	0.39	10	15	0.10	0.00	0.03	1.00	0.95	0.00	0.02
				0.20	0.00	0.02	1.00	0.94	0.00	0.04
			30	0.10	0.00	0.02	1.00	0.94	0.00	0.04
				0.20	0.00	0.03	1.00	0.94	0.00	0.03
		30	15	0.10	0.00	0.02	1.00	0.95	0.00	0.03
				0.20	0.00	0.02	1.00	0.96	0.00	0.02
			30	0.10	0.00	0.03	1.00	0.94	0.00	0.03
				0.20	0.00	0.03	1.00	0.95	0.00	0.03
0.39	0.14	10	15	0.10	0.00	0.00	0.70	0.00	0.30	1.00
				0.20	0.00	0.00	0.71	0.00	0.29	1.00
			30	0.10	0.00	0.00	0.67	0.00	0.33	1.00
				0.20	0.00	0.00	0.70	0.00	0.30	1.00
		30	15	0.10	0.00	0.00	0.65	0.00	0.35	1.00
				0.20	0.00	0.00	0.67	0.00	0.33	1.00
			30	0.10	0.00	0.00	0.65	0.00	0.35	1.00
				0.20	0.00	0.00	0.61	0.00	0.39	1.00
	0.39	10	15	0.10	0.29	1.00	0.10	0.00	0.61	0.00
				0.20	0.31	1.00	0.11	0.00	0.59	0.00
			30	0.10	0.33	1.00	0.11	0.00	0.57	0.00
				0.20	0.30	1.00	0.13	0.00	0.58	0.00
		30	15	0.10	0.25	1.00	0.02	0.00	0.73	0.00
				0.20	0.25	1.00	0.01	0.00	0.74	0.00
			30	0.10	0.21	1.00	0.01	0.00	0.77	0.00
				0.20	0.23	1.00	0.02	0.00	0.75	0.00

Note: "Nonlinear" refers to instantaneous indirect effect estimates using the nonlinear model specification. "Linear" refers to instantaneous indirect effect estimates using the linear model.

Chapter 5: Discussion

The relationship between class size and subsequent student achievement may provide a novel forum within which a nonlinear treatment-mediator relationship may be modeled with the logistic change trajectory. More generally, the logistic change trajectory may provide an appropriate means of modeling nonlinear treatment-mediator relationships containing ceiling and floor effects. The current study was intended to assess estimation of the logistic change trajectory for handling such scenarios.

The current study encountered noticeable estimation problems for two of the investigated models. More specifically, non-convergence rates for the linear outcome model reached values as high as 7.9%, and non-convergences rates for the nonlinear mediator model reached values as high as 44.3%. High non-convergence rates for the linear outcome model stem from conditions with low ICC values, low level-2 sample sizes, and low level-1 sample sizes. The magnitude of the ICC and the number of level-2 units are crucial for proper linear multilevel model estimation; high ICC values and larger numbers of level-2 units are known to produce less biased parameter estimates (Goldstein, 2003), and larger samples are known to facilitate model convergence (Hox, 2002). Given this relationship between ICC and level-2 sample size and multilevel model estimation, the observed higher non-convergence rates for low ICC and low level-2 sample size conditions are unsurprising.

In contrast, the nonlinear mediator model encountered more estimation problems for conditions with larger numbers of level-2 units. This contradicts previous research on

the importance of level-2 sample size in linear multilevel model estimation. For the NLMIXED procedure, larger numbers of random effects (which, in this case, correspond to additional level-2 units) increase the computational complexity of the model's estimation (Kiernan, Tao, & Gibbs, 2012), rendering convergence less likely. In the current investigation, then, conditions with larger numbers of level-2 units should encounter more convergence issues when compared to conditions with lower numbers of level-2 units. Further, additional level-2 random effects create a larger chance that the simulation sample contains outliers. PROC NLMIXED is known to perform poorly with "sufficiently noisy" data (Wolfinger, 1999). As outliers qualify as statistical noise, any conditions that increase the likelihood of their presence will result in increased rates of non-convergence. This was observed in the current study for conditions with larger numbers of level-2 units. Finally, "badly scaled" problems, or problems containing parameters with widely varying scales, introduce computational complexities (SAS, 2012). In the current study, all generating values for the pseudo-slope parameter and its level-2 variance were less than one, while the generating value for the pseudo-intercept parameter was equal to 150. All of these parameters were estimated for each replication, introducing the scaling issue described above. This, combined with conditions containing high level-2 sample sizes, would contribute to increased rates of non-convergence. Scaling issues may be avoided by either a) specifying reasonable starting values for the pseudo-slope parameter, and therefore not requiring that it be estimated, or b) transforming the pseudo-intercept parameter so that its magnitude is closer to that of the pseudo-slope and variance parameters. Future research should explore these possibilities.

Convergence problems for the linear mediator model were only evident in conditions with low level-2 sample sizes and a generating value of 0.39 for the pseudo-slope parameter. However, non-convergence rates for this set of conditions were at most 0.9%, and may therefore be considered negligible. ICC values had no effect on linear mediator model estimation, as the ICC was not considered in generating values for the mediator. Increased non-convergence rates for low level-2 unit conditions mirrors the patterns observed in both linear model estimation and previous research on linear multilevel models (Hox, 2002). For both the linear and nonlinear estimated models for the mediator, non-convergent cases may be explained, in large part, by the large generating value selected for the level-1 residuals' variance. While consistent with the values appearing in the applied example supplied by Singer and Willett (2003), a level-1 variance of 100 may have introduced too much noise into the generated datasets. As PROC NLMIXED is known to encounter convergence issues with noisy data, and problems of scale may affect PROC MIXED convergence (Kiernan, Tao, & Gibbs, 2012), this may explain the observed non-convergence rates for both the linear and the nonlinear models.

The current study used the default settings in SAS PROC NLMIXED when estimating the nonlinear model for the mediator. With non-convergence rates as high as 43%, alternative computational methods may provide more efficient means of estimating the proposed nonlinear model. Additionally, poor starting values for the nonlinear model for the mediator may have contributed to the observed convergence issues. One possible solution may involve estimating the nonlinear model twice (SAS Institute, 2012). In the

first estimation, the first-order optimization method is used in place of the default optimization. Use of the former estimation procedure might provide better starting values that should enhance the resulting convergence rates. Parameters estimates from this method could then be used as starting values for the second nonlinear estimation using the default settings. This method may boost convergence rates in future research.

This study replicates previous findings regarding the recovery of the *b* parameter in multilevel mediation settings (i.e., Pituch et al., 2006). More specifically, this parameter is consistently estimated with bias less than the five percent cutoff recommended by Hoogland and Boomsma (1998) using conventional linear multilevel methods. For conditions containing a true value of zero for the instantaneous indirect effect, both the linear and nonlinear estimators provide unbiased estimates of the instantaneous indirect effect. This is unsurprising given that a) values of zero for the instantaneous indirect effect in this study were produced by specifying a value of zero for the true value of the *b* parameter, and b) the *b* parameter was estimated with minimal bias across all study conditions. For low, mean, and high treatment values, then, unbiased estimation of the *b* parameter leads directly to unbiased estimation of the instantaneous indirect effect when the true value of the *b* parameter is equal to zero.

Estimation of the pseudo-slope parameter returned relative parameter bias values within the recommended cutoff for all but two generating conditions. For these two conditions, however, relative parameter bias of pseudo-slope parameter only marginally exceeded the 0.05 threshold. For the condition in which *b* was generated to be zero, the pseudo-slope parameter value was low, with the smaller level-2 and level-1 sample sizes,

and the lower ICC value, the relative parameter bias for the pseudo-slope was 6.8%. Similarly, for the condition in which *b* was generated to be 0.39, low pseudo-slope, low level-2 and level-1 sample sizes, and the higher ICC value, the relative parameter bias for the pseudo-slope parameter was equal to 5.2%.

The nonlinear estimator proved less biased than the linear estimator at all three treatment levels for all conditions with a non-zero true value for the instantaneous indirect effect. However, only conditions with a pseudo-slope value of 0.39, a *b* parameter value of 0.39, and low treatment values return results within or near Hoogland and Boomsma's (1998) recommended five percent threshold. While the marginal pseudo-slope bias observed for low level-2, low pseudo-slope value conditions necessarily contributed to the observed bias of the instantaneous indirect effect, it in no way accounts for the severity of the bias observed across the majority of the conditions containing a non-zero true instantaneous indirect effect. One possible explanation lies in estimation accuracy of the pseudo-intercept parameter, which was not included as a design condition but was estimated in PROC NLMIXED. Table 9 shows that, for 17 of the 32 design conditions, estimated values for the pseudo-intercept parameter produced relative parameter bias values in excess of 5%. The bias in the estimation of the pseudo-intercept parameter was particularly egregious for conditions with low level-1 and low level-2 sample sizes, ranging from 17% to 22%, and more than doubling the highest parameter bias observed for all other conditions. Because this value goes into the calculation of the instantaneous indirect effect, the biased nature of the pseudo-intercept parameter,

particularly for low level-1 and level-2 sample size conditions, undoubtedly contributed to the bias observed for the instantaneous indirect effect.

Table 9. Relative parameter bias of the pseudo-intercept parameter.

Conditions					
b	γ_{10}	Clusters	Subjects	ICC	Bias
0	0.14	10	15	0.10	**0.162**
				0.20	**0.154**
			30	0.10	**0.059**
				0.20	**0.066**
		30	15	0.10	0.033
				0.20	0.041
			30	0.10	0.019
				0.20	0.012
	0.39	10	15	0.10	**0.170**
				0.20	**0.135**
			30	0.10	**0.074**
				0.20	**0.057**
		30	15	0.10	0.042
				0.20	0.041
			30	0.10	0.022
				0.20	0.015
0.39	0.14	10	15	0.10	**0.156**
				0.20	**0.223**
			30	0.10	**0.070**
				0.20	**0.068**
		30	15	0.10	0.045
				0.20	**0.050**
			30	0.10	0.014
				0.20	0.019
	0.39	10	15	0.10	**0.147**
				0.20	**0.171**
			30	0.10	**0.058**
				0.20	**0.051**
		30	15	0.10	0.043
				0.20	0.038
			30	0.10	0.006
				0.20	0.023

Note. Bolded values indicate relative parameter bias in excess of the recommended 0.05 cutoff.

The magnitude of the true instantaneous indirect effect may also be partially responsible for observed bias results. Given the generating parameter values used in the current investigation, the true instantaneous indirect effect's value was frequently less than one at each treatment value. When conducting the relative parameter bias calculation for certain conditions (as described in Equation 66), this necessarily requires using a value that, in some cases, is substantially less than one for the denominator of the equation. Mathematically, this may result in large relative parameter bias estimates for replications in which the parameter bias is quite negligible. Taking an example from the current study, for the condition with a generating value of 0.39 for both the *b* and pseudo-slope parameters, 10 level-2 units, 30 level-1 units and an ICC of 0.20, the nonlinear estimator produces a relative parameter bias value of 1,872% at high treatment values. The true value of the instantaneous indirect effect for this condition at high treatment values is obtained by substituting the design conditions into Equation 58. Additionally, it can be shown the mean value of a uniform distribution is equal to

$$\frac{\beta + \alpha}{2}, \tag{69}$$

where α and β represent the lower and upper support boundaries of the uniform distribution, respectively (Ross, 2002). Similarly, the standard deviation of a uniform distribution is given by

$$\sqrt{\frac{(\beta - \alpha)^2}{12}}. \tag{70}$$

With a treatment variable drawn from a uniform distribution defined from zero to 50, and the design conditions mentioned above, the true value of the instantaneous indirect effect for high treatment values is equal to 0.000478. For this set of generating conditions, the average value of the instantaneous indirect effect ($\bar{\hat{\theta}}_i$ in the relative parameter bias and parameter bias equations defined Equations 66 and 67) is equal to 0.00942. This produces a parameter bias of 0.00894, but a relative parameter bias of over 1,000%. Table 10 shows that same is true for many of the conditions exhibiting substantial relative parameter bias. It may be, then, that parameter bias rather than relative parameter bias might provide a better metric when the true value of the parameter is substantially less than one. In addition, consideration of the practical significance of the difference between the mean parameter estimate and the true value should inform inferences about how substantial bias might be.

Table 10. Instantaneous indirect effect values for conditions with a non-zero *b* parameter.

Conditions					Treatment Value		
b	γ_{10}	Clusters	Subjects	ICC	Low	Mean	High
0.39	0.14	10	15	0.10	0.15083	0.80884	1.27992
0.39	0.14	10	15	0.20	0.15083	0.80884	1.27992
0.39	0.14	10	30	0.10	0.15083	0.80884	1.27992
0.39	0.14	10	30	0.20	0.15083	0.80884	1.27992
0.39	0.14	30	15	0.10	0.15083	0.80884	1.27992
0.39	0.14	30	15	0.20	0.15083	0.80884	1.27992
0.39	0.14	30	30	0.10	0.15083	0.80884	1.27992
0.39	0.14	30	30	0.20	0.15083	0.80884	1.27992
0.39	0.39	10	15	0.10	3.13907	0.13070	0.00048
0.39	0.39	10	15	0.20	3.13907	0.13070	0.00048
0.39	0.39	10	30	0.10	3.13907	0.13070	0.00048
0.39	0.39	10	30	0.20	3.13907	0.13070	0.00048
0.39	0.39	30	15	0.10	3.13907	0.13070	0.00048
0.39	0.39	30	15	0.20	3.13907	0.13070	0.00048
0.39	0.39	30	30	0.10	3.13907	0.13070	0.00048
0.39	0.39	30	30	0.20	3.13907	0.13070	0.00048

Although PRODCLIN coverage rates for conditions containing true non-zero values of the instantaneous indirect effect favored the nonlinear estimator over the linear estimator, coverage rates for the nonlinear estimator never exceeded 85%, and were as low as 2% for certain combinations of generating parameters and treatment values. Again, this may be a result of the small values for the instantaneous indirect effect under certain conditions. Examining the example from the discussion on the practical significance of relative parameter bias for small true parameter values, the condition with a generating value of 0.39 for both the *b* and pseudo-slope parameters, 10 level-2 units,

30 level-1 units and an ICC of 0.20 resulted in a true instantaneous indirect effect of 0.000478 at high treatment values. Although this value is legitimately positive, estimating a confidence interval around such a negligible effect may be beyond the precision of the estimation methods examined here. Additionally, confidence intervals surrounding such an effect are likely to contain zero within their bounds, leading the applied researcher to conclude that no mediated effect is present. Values this small resulted from choosing generating values for the model parameters and then substituting those into the equation for the instantaneous indirect effect calculation. In future simulation research, it may prove more useful to begin with alternative, though realistic, values for the instantaneous indirect effect at different treatment values, working backwards to derive suitable generating parameters for the nonlinear model, as generating parameter values stemming from different values for the instantaneous indirect effect may be less biased and easier to estimate. Additionally, the observed bias in the nonlinear estimator of the instantaneous indirect effect necessarily decreased the observed confidence interval coverage rates, as coverage rates set around a biased parameter are, themselves, biased. Better estimation of the pseudo-slope and, especially, the pseudo-intercept parameters (as discussed above), may result in improved coverage rates for nonlinear PRODCLIN confidence intervals. Future research should explore how reasonable it is to observe larger values for the instantaneous indirect effect and assess parameter estimation for nonlinear model parameters stemming from practically significant values of the instantaneous indirect effect.

The multivariate delta method that was to calculate a standard error for the nonlinear term for the instantaneous indirect effect utilized the first-order Taylor series expansion of the partial derivative matrix for the nonlinear model parameters. Higher order approximations have been investigated for potential use in linear mediation analysis (e.g., Mackinnon et al, 2002). Although linear multilevel mediation analyses typically utilize the first-order approximation (Mackinnon, 2008), higher order expansion may be warranted in cases involving a nonlinear treatment-mediator relationship. This is especially so for low and high treatment values, as the first-order Taylor series approximation only provides accurate standard errors for functions with independent variables have a high probability of falling near their mean (Oehlert, 1992; Rice, 1994). In the context of mediation analysis, a normally distributed treatment variable would satisfy this requirement, as the treatment value would be most likely to reside near its generating mean. In contrast, the current investigation generated treatment values from a uniform distribution to simulate a dosage-like treatment variable that is both constrained to positive values and that adequately represents the entire spectrum of possible treatment dosages. This latter consideration was especially important for model convergence given the relevant parameter values, as inadequate numbers of observations at the low and high end of the treatment spectrum were shown to severely inhibit model convergence in pilot testing. The uniform distribution investigated here to generate treatment values, then, does not satisfy the condition that treatment values possess a high likelihood of falling near their mean, as values from a uniform distribution are equally distributed across its range of support. In this instance, a higher-order Taylor series expansion may prove

useful when investigating low and high treatment values. Future research should explore this possibility.

Bootstrap resampling methods offer an alternative to the PRODCLIN program for providing confidence intervals around the instantaneous indirect effect (Hayes & Preacher, 2010). A preliminary investigation used the parametric bias-corrected bootstrap resampling method to calculate confidence intervals around the instantaneous indirect effect for one of the conditions (namely, the condition involving generating values of 0.39 for the pseudo-slope parameter, 0.39 for the *b* parameter, 30 level-2 units, 30 level-1 units, and a residual ICC of 0.10) using 1,000 bootstrapped samples per replication. Although only 25 replications were produced, coverage rates were a perfect 100% for low, mean, and high treatment values. However, these improved accuracy over the corresponding PRODCLIN results come at a huge computational cost, as this analysis only produced 25 pilot replications (each with 1,000 bootstrap replications) over a period of 9 days. The added processing time, however, may offset the potential improvement in confidence interval coverage rates. Additionally, the width of the bootstrap confidence interval was larger on average across the 25 bootstrap replications than the PRODCLIN confidence interval for the same set of generating parameters. To the extent that this holds up over a larger number of replications, this may explain why the bootstrap confidence interval outperforms its PRODCLIN counterpart. Future research should extend the research started here and more fully investigate bootstrap coverage rates using a larger number of replications, in addition to examining the trade-off between processing time and coverage rates.

LIMITATIONS

The current investigation is limited in several important ways. First, generating conditions for the pseudo-slope value were limited to values of 0.14 and 0.39. Although the current investigation was designed as an initial attempt to examine the estimation and utility of the logistic change model in the context of multilevel mediation, the importance of the pseudo-slope parameter to both the parameter bias and efficiency of the instantaneous indirect effect suggests a need for the examination of additional generating pseudo-slope values. Further, the generating values for the pseudo-slope parameter were borrowed from the nonlinear mediation literature in the context of dichotomous outcomes. While this provided an adequate starting place for an initial investigation, values drawn from applied examples may prove more useful in examining the multilevel logistic change model. Unfortunately, applied examples of the logistic change trajectory, especially in the context of mediation analysis, are quite sparse at this time. Future research should explore the application of the logistic change model in the context of mediation analysis as a means of exploring plausible generating conditions for subsequent Monte Carlo research.

The current study utilized simulated data based on a model with a fixed pseudo-intercept parameter. This was done for two reasons. First, to mimic the scenario in which participants in a particular intervention must pass a set of global diagnostic or screening criteria. Filtering program participants in this manner ensures that participants are mostly homogenous when beginning the intervention, suggesting a lack of variability in the initial values for the effect of the treatment on a mediator (for a truly mediated construct).

If the screening criteria are well-defined and equally enforced across level-2 units, this lack of pseudo-intercept variability should be evident across all participating clusters. This was modeled in the current study by not allowing the pseudo-intercept parameter to randomly vary across level-2 units. Secondly, this particular model specification (without a random effect for the pseudo-intercept parameter) was chosen to facilitate model convergence, as SAS PROC NLMIXED is optimized for problems with a single random effect (Wolfinger, 1999). Substantively, the former consideration may not hold across all interventions that utilize a prescreening process. Mathematically, SAS PROC NLMIXED is capable of estimating models containing multiple random effects, within the limits of the scaling and level-2 sample size considerations mentioned previously. Future research should explore scenarios in which both the pseudo-intercept and the pseudo-slope parameters are allowed to vary across level-2 units.

The generating value of the pseudo-intercept parameter was not included as a design variable and was, instead, set at a value of 150 for all study conditions. This value was presented in Singer and Willett's (2003) illustration, as were other plausible values for the pseudo-intercept parameter. The *a posteriori* bias analysis for the pseudo-intercept parameter (presented above) suggests that estimation issues surrounding this parameter contribute to the observed relative parameter bias of the instantaneous indirect effect. Given the paucity of applied examples utilizing the logistic change trajectory, additional research is needed to understand the range of appropriate generating values for future simulation research. As an additional consideration, the scale of this value was noticeably different from the scale of the pseudo-slope parameter. Given the previously discussed

scaling considerations relevant to models estimated with SAS PROC NLMIXED, this may have contributed to the observed convergence issues. Future research should explore the possibility of transforming the pseudo-intercept variable in estimating the logistic change trajectory.

The upper and lower asymptotes for the nonlinear mediator model were not estimated by SAS PROC NLMIXED and were, instead, explicitly specified as part of the estimation process. This is not a requirement of the logistic change trajectory or the estimation process, as these values may be estimated by various statistical optimization procedures (PROC NLMIXED included). However, given the research on similar model specifications, the upper and lower asymptotes may prove difficult to estimate. More specifically, the similarity between the three-parameter logistic (3PL) model described in the item response theory (IRT) literature and the logistic change trajectory investigated here may provide some insight into the estimation of the upper and lower asymptotes. Briefly, the 3PL model can be specified as a generalized linear model whereby the probability of a correct response to item *i* for subject *j* is given according to the following equation:

$$P(X_{ij}=1|\theta_j,a_i,b_i,c_i)=c_i+\frac{(1-c_i)}{1+\exp[-1.7a_i(\theta_j-b_i)]}. \quad (71)$$

Here, X_{ij} represents the response to item *i* for person *j* (coded 1 if the item is answered correctly, and zero otherwise), θ_j represents the latent ability of person *j*, a_i represents the discrimination parameter for item *i*, b_i represents the model estimated difficulty

parameter for item i, and c_i represents the pseudo-guessing parameter for item i (Birnbaum, 1968). The logit link function is used to specify the relationship between the observed response and the model's parameters, resulting in a model that may be estimated as a multilevel logistic regression model with items (level-1) nested within subjects (level-2) an ability parameter that varies across level-2 units. The 3PL model may be considered a special case of the logistic change trajectory whereby additional parameters (specifically θ_j, a_i, and b_i) are included within the exponent, the pseudo-intercept and upper asymptote parameters are set equal to 1, and the lower asymptote, c_i, is estimated from the data. Given this, research regarding estimation of the lower asymptote in the 3PL model may inform potential issues regarding the estimation of the both asymptote parameters in the logistic change trajectory. More notably, the pseudo-guessing parameter is particularly difficult to estimate, requiring larger sample sizes and, in many cases, Bayesian estimation methods (de Ayala, 2009). Similar problems are likely to be encountered when estimating the logistic change trajectory's asymptote parameters. Future researchers should be wary of these potential complications and explore alternatives for enhancing the asymptotes' estimation.

This study included only one non-zero generating value for the *b* parameter. For conditions with a generating value of zero for the *b* parameter, relative parameter bias values for the instantaneous indirect effect estimates were equal to zero (to two decimal places) across all other study conditions at low, mean, and high treatment values. In contrast, the majority of the conditions with a value of 0.39 for the *b* parameter exhibited

substantial relative parameter bias. The relationship between the magnitude of the *b* parameter and the extent of the relative parameter bias, however, was impossible to determine given the current set of design conditions. Additionally, only positive values of the *b* parameter were considered. Future research should explore the relationship between both the magnitude and the sign of the *b* parameter and the resulting relative parameter bias.

Although the current investigation provides nonlinear specifications for $1 \rightarrow 1 \rightarrow 1$, $2 \rightarrow 1 \rightarrow 1$, and $2 \rightarrow 2 \rightarrow 1$ designs, this simulation study only examined the statistical performance of the $1 \rightarrow 1 \rightarrow 1$ specification. Future research should focus on the $2 \rightarrow 1 \rightarrow 1$ and $2 \rightarrow 2 \rightarrow 1$ designs that are commonly encountered in cluster randomized trials. In addition, the proposed study only examined multilevel designs containing at most two-levels consisting of a nonlinear relationship between the treatment and the mediating variables. Nonlinear parameterizations could easily be extended to 3-level designs (e.g., Pituch, Murphy, & Tate, 2010) or to designs specifying nonlinearity between the mediating and the outcome variables. Future research should examine these possibilities. Nonlinear multilevel mediation may also exist in the context of repeated measures or longitudinal data analysis. In these instances, outcomes scores are measured within individuals over time. Singer and Willett (2003) suggest that these kinds of designs may lend themselves to nonlinear model specifications. Future research should explore these nonlinear longitudinal designs in the context of mediation analysis.

The current investigation examined one specific nonlinear treatment-mediator relationship. As previously mentioned, nonlinear relationships are not limited to the

logistic change trajectory examined here. Polynomial and other exponential relationships may provide better model fit in certain circumstances. Additionally, although the nonlinear estimator outperformed the linear estimator for the majority of generating conditions, this may not be true of other nonlinear specifications. For complex nonlinear models with multiple parameters to be estimated, the complexity of the estimation process may render linear methods more desirable. Other nonlinear models will perform differently than the current nonlinear specification, and the results presented here may not generalize. However, although the current investigation is model-specific, the procedure described here does generalize to other mediation analyses that specify a nonlinear treatment-mediator relationship. This procedure is summarized as follows:

1. Specify a model for the treatment-mediator relationship and for outcome as a function of both the treatment and the mediator.
2. Using the mathematical definition of the instantaneous indirect effect, derive the indirect effect from the equations for the mediator and the outcome.
3. Using the delta method, derive the standard error for the nonlinear component of the instantaneous indirect effect. If any portion of the mediated relationship is specified as linear, parameter standard errors will be provided as part of the estimation process.
4. Estimate all necessary parameters.
5. Substitute the estimated parameters into the equation for the instantaneous indirect effect derived in 2. If the resulting instantaneous indirect effect is a function of the treatment variable, use three treatment values in calculating the

instantaneous indirect effect: the sample mean treatment value, and values one standard deviation above and below the sample mean treatment value.

6. Supply values for the instantaneous indirect and all relevant standard errors to PRODCLIN to estimate a confidence interval at each of the three treatment values. Confidence intervals containing zero indicate that the treatment effect is not mediated at the current treatment value.

These steps provide a general outline for incorporating a nonlinear relationship into a mediation analysis.

In this study, the generating model for the treatment-mediator relationship was known to be nonlinear in nature. In reality, the relationship underlying a set of observations is rarely known *a priori*. The utility of specifying a nonlinear relationship for any portion of a mediation pathway rests upon the assumption that the relationship in question is truly nonlinear. Testing the nonlinearity of an observed relationship may require a combination of quantitative and qualitative analysis. Quantitatively, comparison of model fit indices and standardized residuals across linear and nonlinear model specifications may provide applied researchers with some guidance during the model selection process. Content experts and theoretical frameworks may provide additional information where analytic endeavors provide ambiguous answers. Applied researchers should be aware of these considerations prior to implementing the steps described above.

Finally, although the proposed nonlinear multilevel mediation parameterization has roots in latent variable analysis, the current investigation did not utilize structural equation modeling methods to investigate the instantaneous indirect effect. Previous

research has examined structural equation modeling as a means of providing confidence intervals around the mediated effect (Cheung, 2007; Bollen & Stine, 1990), and as a means of providing alternate specifications for meditational models (Winship & Mare, 1983). Further, the conflation of within- and between- group variance, as discussed previously, may be best addressed in the context of MSEM. Future research should explore the possibility of using MSEM to specify nonlinear mediating relationships. Additionally, multiple mediating measures may be employed as observed indicators of an unobserved latent mediating variable. This latent mediator may also be nonlinearly related to the treatment variable, resulting in a nonlinear mediation model similar to the model proposed in the current investigation. This would then necessitate estimation of a multilevel structural equation model to properly address the clustering of individuals within level-two units. Again, this is only one example of many possible alternative parameterizations. Future research should investigate the performance of nonlinear multilevel latent variable models in the context of meditational analysis.

Overall, the nonlinear estimator outperformed the linear estimator in terms of bias, efficiency, and PRODCLIN coverage rates. Generating values for the pseudo-slope parameter and the number of level-2 units had the most noticeable effect on convergence, bias, efficiency, and PRODCLIN coverage rates, suggesting the need for further simulations using additional generating values for these parameters. Moreover, the overall superior performance of the nonlinear estimator suggests that linear approximations for mediation analyses may not always be viable in the context of truly nonlinear relationships. While the parsimony of linearity is appealing for both practical

and philosophical reasons, linear methods do not provide an analytical panacea, especially in the context of mediation analysis. To the extent that both theory and analytical methods support nonlinear specifications, researchers should consider alternatives to conventional linear mediation model specifications.

Appendix

Derivation of $\frac{\partial M}{\partial T}$:

$$\frac{\partial M}{\partial T} = \frac{\partial}{\partial T}\left(\alpha_1 + \frac{(\alpha_2 - \alpha_1)}{1 + \gamma_{00}\exp[-(\gamma_{10}T)]}\right)$$

1) Differentiate each term in the sum; factor out constants.

$$= (\alpha_2 - \alpha_1)\left\{\frac{\partial}{\partial T}\left(\frac{1}{1 + \gamma_{00}\exp[-(\gamma_{10}T)]}\right)\right\} + \frac{\partial}{\partial T}(\alpha_1)$$

2) The derivative of α_1, a constant, is zero.

$$= (\alpha_2 - \alpha_1)\left\{\frac{\partial}{\partial T}\left(\frac{1}{1 + \gamma_{00}\exp[-(\gamma_{10}T)]}\right)\right\} + 0$$

3) Chain rule: $\frac{\partial M}{\partial T} = \left(\frac{\partial}{\partial u}\cdot\frac{1}{u}\right)\cdot\frac{\partial u}{\partial T}$, with $u = 1 + \gamma_{00}\exp[-(\gamma_{10}T)]$, $M = \frac{1}{u}$, and $\frac{\partial}{\partial u}\cdot\frac{1}{u} = -\frac{1}{u^2}$.

$$(\alpha_2 - \alpha_1)\left\{\frac{\partial}{\partial T}\left(\frac{1}{1 + \gamma_{00}\exp[-(\gamma_{10}T)]}\right)\right\} = (\alpha_2 - \alpha_1)\left(\frac{\partial}{\partial u}\cdot\frac{1}{u}\right)\cdot\frac{\partial u}{\partial T} = -(\alpha_2 - \alpha_1)\left(\frac{1}{u^2}\cdot\frac{\partial u}{\partial T}\right)$$

Substituting for *u*:

$$(\alpha_2 - \alpha_1)\left(-\frac{1}{u^2}\cdot\frac{\partial u}{\partial T}\right) = -(\alpha_2 - \alpha_1)\left\{\frac{\frac{\partial}{\partial T}\{1 + \gamma_{00}\exp[-(\gamma_{10}T)]\}}{\{1 + \gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right\}$$

4) Differentiate the numerator sum by term and factor out constants.

$$-(\alpha_2-\alpha_1)\frac{\frac{\partial}{\partial T}\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}=-(\alpha_2-\alpha_1)\left\{\frac{\frac{\partial}{\partial T}(1)+\left\{\gamma_{00}\left[\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)])\right]\right\}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right\}$$

The derivative of one, a constant, is zero;

$$=-(\alpha_2-\alpha_1)\left\{\frac{0+\left\{\gamma_{00}\left[\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)])\right]\right\}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right\}$$

5) Chain rule: $\frac{\partial \exp[-(\gamma_{10}T)]}{\partial T}=\frac{\partial \exp[u]}{\partial u}\cdot\frac{\partial u}{\partial T}$, where $u=-\gamma_{10}T$ and $\frac{\partial \exp[u]}{du}=\exp[u]$.

$$-(\alpha_2-\alpha_1)\left\{\frac{\left\{\gamma_{00}\left[\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)])\right]\right\}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right\}=-\gamma_{00}(\alpha_2-\alpha_1)\frac{\exp[-(\gamma_{10}T)]\left\{-\left[\frac{\partial}{\partial T}(\gamma_{10}T)\right]\right\}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}$$

6) Factor out constants:

$$-\gamma_{00}(\alpha_2-\alpha_1)\frac{\exp[-(\gamma_{10}T)]\left\{-\left[\frac{\partial}{\partial T}(\gamma_{10}T)\right]\right\}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}=\gamma_{00}\gamma_{10}(\alpha_2-\alpha_1)\frac{\exp[-(\gamma_1T)]\left\{\left[\frac{\partial}{\partial T}(T)\right]\right\}}{\{1+\gamma_{00}\exp[-(\gamma_1T)]\}^2}$$

7) The derivative of T is 1.

$$\gamma_{00}\gamma_{10}(\alpha_2-\alpha_1)\frac{\exp[-(\gamma_{10}T)]\left\{\left[\frac{\partial}{\partial T}(T)\right]\right\}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}=\frac{\gamma_{00}\gamma_{10}(\alpha_2-\alpha_1)\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}$$

Derivation of $\frac{\partial \theta}{\partial T}$:

$$\frac{\partial \theta}{\partial T} = \frac{\partial}{\partial T}\left(\frac{\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} b \right)$$

1) Factor out constants:

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\frac{\partial}{\partial T}\left(\frac{\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} \right)$$

2) Product rule: $\frac{\partial}{\partial T}(u \cdot v) = v\frac{\partial u}{\partial T} + u\frac{\partial v}{\partial T}$, where $u = \exp[-(\gamma_{10}T)]$ and $v = \frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}$.

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(\frac{\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)])}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} + \exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}\left(\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} \right) \right) \right)$$

3) Chain rule: $\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)]) = \frac{\partial e^u}{\partial u} \cdot \frac{\partial u}{\partial T}$, where $u = -\gamma_{10}T$ and $\frac{\partial e^u}{\partial u} = e^u$.

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(\frac{\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}(-\gamma_{10}T) \right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} + \exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}\left(\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} \right) \right) \right)$$

4) Chain rule: $\frac{\partial}{\partial T}\left(\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} \right) = \frac{\partial}{\partial u} \cdot \frac{1}{u^2} \cdot \frac{\partial u}{\partial T}$, where $u = \gamma_{00}\exp[-(\gamma_{10}T)]$ and $\frac{\partial}{\partial u} \cdot \frac{1}{u^2} = \frac{-2}{u^3}$.

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(\frac{\exp[-(\gamma_{10}T)]\frac{\partial}{\partial T}(-\gamma_{10}T)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} - \exp[-(\gamma_{10}T)]\left(\frac{2\left(\frac{\partial}{\partial T}(\gamma_{00}\exp[-(\gamma_{10}T)]+1) \right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3} \right) \right)$$

5) Factor out constants:

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(\left(-\frac{\gamma_{10}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}(T)\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right) - \frac{2\gamma_{00}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)]+1)\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)$$

6) Differentiate each component of the second term's numerator:

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(-\frac{\gamma_{10}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}(T)\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} - \frac{2\gamma_{00}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}\left(\exp[-(\gamma_{10}T)]\right)+\frac{\partial}{\partial T}(1)\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)$$

7) $\frac{\partial}{\partial T}(T) = 1$, $\frac{\partial}{\partial T}(1) = 0$.

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(-\frac{\gamma_{10}\exp[-(\gamma_{10}T)](1)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} - \frac{2\gamma_{00}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)])+0\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)$$

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(-\frac{\gamma_{10}\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} - \frac{2\gamma_{00}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)])\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)$$

8) Chain rule: $\frac{\partial}{\partial T}(\exp[-(\gamma_{10}T)]) = \frac{\partial e^u}{\partial u}\cdot\frac{\partial u}{\partial T}$, where $u = -\gamma_{10}T$ and $\frac{\partial e^u}{du} = e^u$.

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(-\frac{\gamma_{10}\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} - \frac{2\gamma_{00}\exp[-(\gamma_{10}T)]\left((\exp[-(\gamma_{10}T)])\frac{\partial}{\partial T}(-\gamma_{10}T)\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)$$

9) Factor out constants:

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(-\frac{\gamma_{10}\exp[-(\gamma_{10}T)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} - \frac{2\gamma_{00}\exp[-(2\gamma_{10}T)]\left(-\gamma_{10}\frac{\partial}{\partial T}(T)\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)$$

10) The derivative of T is 1.

$$= [\gamma_{00}\gamma_{10}(\alpha_2 - \alpha_1)b]\left(\frac{2\gamma_{00}\gamma_{10}\exp[-(2\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3} - \frac{\gamma_{10}\exp[-(\gamma_{10}T)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2} \right)$$

Derivation of the multivariate delta standard error for $\left\{\frac{\gamma_{00}\gamma_{10}(\alpha_2-\alpha_1)\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right\}$:

SE = D'VD, where

V = $\begin{bmatrix} \sigma^2_{\gamma_{00}} & \sigma_{\gamma_{00}\gamma_{10}} \\ \sigma_{\gamma_{00}\gamma_{10}} & \sigma^2_{\gamma_{00}} \end{bmatrix}$ from the mediator model estimation, and

D = $\begin{bmatrix} \frac{\partial Y}{\partial \gamma_{00}} \\ \frac{\partial Y}{\partial \gamma_{10}} \end{bmatrix}$, where Y = $\left\{\frac{\gamma_{00}\gamma_{10}(\alpha_2-\alpha_1)\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right\}$.

Evaluate $\frac{\partial Y}{\partial \gamma_{00}}$.

1) Factor out constants.

$$\frac{\partial Y}{\partial \gamma_{00}} = (\alpha_2-\alpha_1)\gamma_{10}\exp[(-\gamma_{10}T)]\frac{\partial}{\partial \gamma_{00}}\left(\frac{\gamma_{00}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right)$$

2)Product rule: $\frac{\partial}{\partial \gamma_{00}}(uv) = v\frac{\partial u}{\partial \gamma_{00}} + u\frac{\partial v}{\partial \gamma_{00}}$, where $u = \gamma_{00}$ and $v = \frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}$.

$$= (\alpha_2-\alpha_1)\gamma_{10}\exp[-(\gamma_{10}T)]\left(\gamma_{00}\left(\frac{\partial}{\partial \gamma_{00}}\left(\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right)\right) + \frac{\frac{\partial(\gamma_{00})}{\partial \gamma_{00}}}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right)$$

3) $\frac{\partial(\gamma_{00})}{\partial \gamma_{00}} = 1$.

$$= (\alpha_2 - \alpha_1)\gamma_{10}\exp[-(\gamma_{10}T)]\left(\gamma_{00}\left(\frac{\partial}{\partial\gamma_{00}}\left(\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right)\right)+\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right)$$

4) Chain rule: $\frac{\partial}{\partial\gamma_{00}}\left(\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right) = -\frac{2\frac{\partial u}{\partial\gamma_{00}}}{u^3}$, where and $\frac{\partial\frac{1}{u^2}}{\partial u} = -\frac{2}{u^3}$.

=

$$(\alpha_2 - \alpha_1)\gamma_{10}\exp[-(\gamma_{10}T)]\left(\gamma_{00}\left(-\left(\frac{2\left(\frac{\partial}{\partial\gamma_{00}}\left(1+\gamma_{00}\exp[-(\gamma_{10}T)]\right)\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)\right)+\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right)$$

5) Differentiate the sum by term and factor out constants.

$$= (\alpha_2 - \alpha_1)\gamma_{10}\exp[-(\gamma_{10}T)]\left(\left(-\left(\frac{2\gamma_{00}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial\gamma_{00}}(\gamma_{00})+\frac{\partial}{\partial\gamma_{00}}(1)\right)}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)\right)+\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}\right)$$

6) Evaluate remaining derivatives.

$$= (\alpha_2 - \alpha_1)\gamma_{10}\exp[-(\gamma_{10}T)]\left(\frac{1}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^2}-\frac{2\gamma_{00}\exp[-(\gamma_{10}T)]}{\{1+\gamma_{00}\exp[-(\gamma_{10}T)]\}^3}\right)$$

Evaluate $\frac{\partial Y}{\partial \gamma_{10}}$.

1) Factor out constants.

$$\frac{\partial Y}{\partial \gamma_{10}} = (\alpha_2 - \alpha_1)\gamma_{00} \frac{\partial}{\partial \gamma_{00}} \left(\frac{\gamma_{10} \exp[(-\gamma_{10} T)]}{\{1 + \gamma_{00} \exp[-(\gamma_{10} T)]\}^2} \right)$$

2) Product rule: $\frac{\partial}{\partial \gamma_{00}}(uv) = v\frac{\partial u}{\partial \gamma_{10}} + u\frac{\partial v}{\partial \gamma_{10}}$, where $u = \exp[-(\gamma_{10} T)]$ and $v = \frac{\gamma_{10}}{(1 + \gamma_{00} \exp[-(\gamma_{10} T)])^2}$.

$$= (\alpha_2 - \alpha_1)\gamma_{00} \left(\exp[-(\gamma_{10} T)] \left(\frac{\partial}{\partial \gamma_{10}} \left(\frac{\gamma_{10}}{(1 + \gamma_{00} \exp[-(\gamma_{10} T)])^2} \right) \right) + \frac{\gamma_{10} \left(\frac{\partial}{\partial \gamma_{10}} (\exp[-(\gamma_{10} T)]) \right)}{(1 + \gamma_{00} \exp[-(\gamma_{10} T)])^2} \right)$$

3) Chain rule: $\frac{\partial}{\partial \gamma_{10}} (\exp[-(\gamma_{10} T)]) = \exp(u)\frac{\partial u}{\partial \gamma_{10}}$, where $u = -\gamma_{10} T$ and $\frac{\partial e}{\partial u} = \exp(u)$.

$$= (\alpha_2 - \alpha_1)\gamma_{00} \left(\gamma_{10} \exp[-(\gamma_{10} T)] \left(\frac{\frac{\partial}{\partial \gamma_{10}}(-\gamma_{10} T)}{(1 + \gamma_{00} \exp[-(\gamma_{10} T)])^2} \right) + (\exp[-(\gamma_{10} T)]) \left(\frac{\partial}{\partial \gamma_{10}} \left(\frac{\gamma_{10}}{(1 + \gamma_{00} \exp[-(\gamma_{10} T)])^2} \right) \right) \right)$$

4) Product rule: $\frac{\partial}{\partial \gamma_{00}}(uv) = v\frac{\partial u}{\partial \gamma_{10}} + u\frac{\partial v}{\partial \gamma_{10}}$, where $u = \frac{1}{(1 + \gamma_{00} \exp[-(\gamma_{10} T)])^2}$ and $v = \gamma_{10}$.

$$= (\alpha_2 - \alpha_1)\gamma_{00} \left(\exp[-(\gamma_{10} T)] \left(\gamma_{10} \left(\frac{\partial}{\partial \gamma_{10}} \left(\frac{1}{1 + \gamma_{00} \exp[-(\gamma_{10} T)]^2} \right) \right) + \frac{\frac{\partial}{\partial \gamma_{10}}(\gamma_{10})}{(1 + \gamma_{00} \exp[-(\gamma_{10} T)])^2} \right) + \frac{\gamma_{10} \exp[-(\gamma_{10} T)] \left(\frac{\partial}{\partial \gamma_{10}}(-\gamma_{10} T) \right)}{(1 + \gamma_{00} \exp[-(\gamma_{10} T)])^2} \right)$$

5) Factor out constants:

$$= (\alpha_2 - \alpha_1)\gamma_{00}\left(\exp[-(\gamma_{10}T)]\left(\gamma_{10}\left(\frac{\partial}{\partial \gamma_{10}}\left(\frac{1}{1+\gamma_{00}\exp[-(\gamma_{10}T)]^2} \right) \right) + \frac{\frac{\partial}{\partial \gamma_{10}}(\gamma_{10})}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} \right) - \frac{\gamma_{10}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial \gamma_{10}}(\gamma_{10}T) \right)}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} \right)$$

6) Chain rule: $\frac{\partial}{\partial \gamma_{10}}\left(\frac{1}{1+\gamma_{00}\exp[-(\gamma_{10}T)]^2} \right) = -\frac{2\frac{\partial u}{\partial \gamma_{10}}}{u^3}$, where $u = 1+\gamma_{00}\exp[-(\gamma_{10}T)]$ and $\frac{\partial \frac{1}{u^2}}{\partial u} = -\frac{2}{u^3}$:

$$= (\alpha_2 - \alpha_1)\gamma_{00}\left(\exp[-(\gamma_{10}T)]\left(\frac{\frac{\partial}{\partial \gamma_{10}}(\gamma_{10})}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} + \gamma_{00}\left(-\frac{2\left(\frac{\partial}{\partial \gamma_{10}}(1+\gamma_{00}\exp[-(\gamma_{10}T)]) \right)}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^3} \right) \right) - \frac{\gamma_{10}\exp[-(\gamma_{10}T)]\left(\frac{\partial}{\partial \gamma_{10}}(\gamma_{10}T) \right)}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} \right)$$

7) Differentiate the sum by term and factor out constants:

$$= (\alpha_2 - \alpha_1)\gamma_{00}\left(\exp[-(\gamma_{10}T)]\left(\frac{1}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} - \left(\frac{2\gamma_{00}\gamma_{10}\left(\frac{\partial}{\partial \gamma_{10}}(\exp[-(\gamma_{10}T)]) \right)}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^3} \right) \right) - \frac{\gamma_{10}T\exp[-(\gamma_{10}T)]}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} \right)$$

8) Chain rule: $\frac{\partial}{\partial \gamma_{10}}(\exp[-(\gamma_{10}T)]) = \exp(u)\frac{\partial u}{d\gamma_{10}}$, where $u = -\gamma_{10}T$ and $\frac{\partial \exp(u)}{du} = \exp(u)$.

$$= (\alpha_2 - \alpha_1)\gamma_{00}\left(\exp[-(\gamma_{10}T)]\left(\frac{1}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} - \left(\frac{2\gamma_{00}\gamma_{10}\left(\exp[-(\gamma_{10}T)]\frac{\partial}{\partial \gamma_{10}}(-\gamma_{10}T) \right)}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^3} \right) \right) - \frac{\gamma_{10}T\exp[-(\gamma_{10}T)]}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} \right)$$

9) Factor out constants and complete remaining derivatives:

$$= (\alpha_2 - \alpha_1)\gamma_{00}\left(\exp[-(\gamma_{10}T)]\left(\frac{1}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} + \left(\frac{2\gamma_{00}\gamma_{10}T(\exp[-(\gamma_{10}T)])}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^3} \right) \right) - \frac{\gamma_{10}T\exp[-(\gamma_{10}T)]}{(1+\gamma_{00}\exp[-(\gamma_{10}T)])^2} \right)$$

References

Aroian, L. A. (1944). The probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics, 18*, 265-271.

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Education Statistics, 6*, 267-285.

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.

Bauer, D. J., & Cai, L. (2009). Consequences of unmodeled nonlinear effects in multilevel models. *Journal of Educational and Behavioral Statistics, 34(1)*, 97-114.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychological Methods, 11*(2), 142-163.

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: classical and bootstrap estimates of variability. *Sociological Methodology, 20*, 115-140.

Bruner, J. S. (1957). Going beyond the information given. In H. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary approaches to cognition* (pp. 258-290). Cambridge, MA: Harvard University Press.

Collins, L. M., Graham, J. W., & Flaherty, B. P. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research, 33*, 95-312.

Cheung, M. W.-L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling,14(2)*, 227-246.

Craig, C. C. (1936). On the frequency function of xy. *Annals of Mathematical Statistics, 7*, 1-15.

Cudeck, R., & du Toit, S. H. C. (2002). A nonlinear form of quadratic regression with interpretable parameters. *Multivariate Behavioral Reserch, 37*, 501-519.

Cudeck, R. & Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology, 58*, 615-637.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: The Guillford Press.

Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium.* New Haven, Conn.: Yale University Press.

Finn, J. D., and C. M. Achilles. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*, 557-577.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51*, 115-134.

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

Goldstein, H. (2003). *Multilevel statistical models (3rd ed)*. London: Edward Arnold.

Hanushek, E. A., & Jackson, J. E. (1977). *Statistical Methods for Social Scientists.* New York: Academic Press.

Hayes, A. F., & Preacher, K. J. (2010). Quantifying and testing indirect effects in simple mediation models when the constituent paths are nonlinear. *Multivariate Behavioral Statistics, 45*, 627-660.

Homer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression. 2nd Ed.* New York: John Wiley and Sons.

Hoogland, J.J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26*, 329–367.

Howell, D. S. (2002). *Statistical Methods for Psychology. 5th Ed.* Belmont, CA: Wadsworth.

Hox, J. J. (2002). Multilevel analysis: Techniques and applications. Mahwah, NJ: Erlbaum.

Jasti, S., Dudley, W. N., & Goldwater, E. (2008). SAS macros for testing statistical mediation in data with binary mediators or outcomes. *Nursing Research, 57*(2), 118-122.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: estimating mediation in treatment evaluations. *Evaluation Review, 5*, 602-619.

Johnson, N., & Kotz, S. (1995). *Continuous Univariate Distributions (Volume 1). 2nd Ed.* New York: Wiley.

Kahneman, D., & Tversky, A. (1979).Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263-291.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.). *The handbook of social psychology* (4th ed., Vol. 1, pp. 223-265). New York: McGraw-Hill.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods, 8*(2), 115-128.

Kiernan, K., Tao, J., & Gibbs, P. (2012). SAS white paper on tips and strategies for mixed modeling with SAS/STAT procedures [White paper]. Retrieved from http://support.sas.com/resources/papers/proceedings12/332-2012.pdf

Klein, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.

Knobloch,, L. K. (2007). Perceptions of turmoil within courtship: Associations with intimacy, relational uncertainty, and interference from partners. *Personal Relationships, 14*, 91-112.

Konstantopoulos, S. (2011). How consistent are class size effects? *Evaluation Review, 35*, 71-92.

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from project STAR. *Elementary School Journal, 108*, 275-291.

Kreft, I. G. G. (1996). *Are multilevel techniques necessary? An overview, including simulation studies.* Retrieved December 5, 2010, from http://www.calstatela.edu/faculty/ikreft/quarterly/quarterly.html.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114*, 497-532.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel modeling in group-based intervention studies. *Evaluation Review, 23*, 418-444.

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

Li, Y., Schneider, J. A., & Bennett, D. A. (2007). Estimation of the mediation effect with a binary mediator. *Statistics in Medicine, 26*, 3398-3414.

Lomnicki, Z. A. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society, 29*, 513-524.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*(2), 144-158.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007a). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods, 39*(3), 384-389.

MacKinnon, D. P., Lockwood, C. M., Brown, C. H., Wang, W., & Hoffman, J. M. (2007b). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials, 4*(5), 499-513.

MacKinnon, D. P., Lockwood, J. L., 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.

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

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effects. *Prevention Science, 1*, 173-181.

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. New York: Lawrence Erlbaum Associates.

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

Mehta, P. D., & Neale, M. C. (2005) People are variables too: multilevel structural equations modeling. *Psychological Methods, 10(3),* 259-284.

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39*(1), 129-149.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics, 32*, 385-397.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267-316.

Neuhaus, J., & McCulloch, C. (2006). Separating between- and within-cluster covariate effects by using conditional partitioning methods. *Journal of the royal Statistical Society, Series B: Statistical Methodology, 68*, 859-872.

Oehlert, G. W. (1992). A note on the delta method. *American Statistician (46)*, 27–29.

Pituch, K. A., & Miller, J. W. (1999). Strengthening multisite educational interventions: An illustration with multilevel modeling. *Educational Research and Evaluation, 5*, 62-75.

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

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

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*, 237-267.

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

Plewis, I., & Hurry, J. (1998). A multilevel perspective on the design and analysis of intervention studies. *Educational Research and Evaluation, 4*, 13-26.

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*(3), 879-891.

Preacher, K. J. (2011). Multilevel SEM strategies for evaluating mediation in three-level data. *Multivariate Behavioral Research, 46(4)*, 694-731.

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

Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods. 2nd Ed.* Newbury Park, CA: Sage.

Rice, John. (1994). *Mathematical Statistics and Data Analysis*. *2nd ed*. Duxbury.

Robertson, T. B. (1909). A biochemical conception of memory and sensation. *The Monist, 19*, 367-386.

Ross, S. (2002). *A first course in probability. 6th edition.* Upper Saddle River, NJ: Prentice-Hall.

SAS Institute (2009). *User's Guide: Statistics, Version 9.2. 2nd Ed.* Cary, NC: Author.

SAS Institute (2012). Computational Problems. In *SAS 9.22 User's Guide*. Retrieved October 16, 2012, from http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_nlmixed_sect032.htm

Scariano, S. M., & Davenport, J. M. (1987). The effects of violations of independence assumptions in one-way ANOVA. *The American Statistician, 41*, 123-129.

Scott, A. J., & Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association, 77*, 848-854.

Shin, Y., & Raudenbush, S. W. (2011). The causal effect of class size on academic achievement: Multivariate instrumental variable estimars with data missing at random. *Journal of Educational and Behavioral Statistics, 36*, 154-185.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422-445.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York: Oxford University Press.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290-312). San Francisco: Jossey-Bass.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. Tuma (Ed.), *Sociological methodology* (pp. 159-186). Washington, DC: American Sociological Association.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than meditational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*, 845-851.

Springer, M. D., & Thompson, W. E. (1966). The distribution of independent random variables. *SIAM Journal on Applied Mathematics, 14*, 511-526.

Stolzenberg, R. M. (1980). The measurement and decomposition of causal effects in nonlinear and nonadditive models. *Sociological Methodology, 11*, 459-488.

Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika, 55*, 337-352.

Walsh, J. E. (1947). Concerning the effect of the intraclass correlation on certain significance tests. *Annals of Mathematical Statistics, 18*, 88-96.

Winship, C., & Mare, R. D. (1983). Structural equations and path analysis for discrete data. *American Journal of Sociology, 89*, 54-110.

Wolfinger, R.D. (1999). Fitting nonlinear mixed models with the new NLMIXED procedure. Proceedings of the Twenty-Fourth Annual SAS Users Group Conference, SAS Institute Inc., Cary, NC.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482.