**Abstract**

Mediation analysis is a statistical approach used to examine how the effect of an independent variable on an outcome is transmitted through an intervening variable (mediator). In this article, we provide a gentle introduction to single-level and multilevel mediation analyses. Using single-level data, we demonstrate an application of structural equation modeling (SEM) in estimating mediation models with multiple mediators and multiple outcomes. We also describe the estimation and interpretation of a $2 \rightarrow 1 \rightarrow 1$ multilevel mediation model in which the first, second, and third numbers correspond to the measurement levels of the independent, mediator, and outcome variables, respectively. We present two numerical examples that use simulated data based on published studies in adolescence research to demonstrate how to specify, estimate, and interpret the results of single-level and multilevel mediation analyses to answer key research questions. We briefly discuss underlying assumptions required to make a valid causal interpretation of a mediation analysis.

**Keywords**

mediation analysis, multilevel mediation analysis
Mediation analysis is a statistical approach used to understand how the effect of an independent variable on an outcome is transmitted through an intervening variable, commonly referred to as a mediator. Conceptually, the mediated (indirect) effect is the effect of the independent variable on the outcome through the mediator. The effect of the independent variable on the outcome that is not mediated is called a direct effect. Many examples of mediation analysis can be found in early adolescence research (e.g., Hektner & Swenson, 2012; Loukas, Suzuki, & Horton, 2006; Torsheim & Wold, 2001; Vieno, Santinello, Pastore, & Perkins, 2007).

The classic mediation model (that assumes independent observations, that is, no clustering) has been extensively studied (Baron & Kenny, 1986; James & Brett, 1984; Preacher & Hayes, 2004; Tofighi, MacKinnon, & Yoon, 2009). Extending the classic mediation model to clustered data (e.g., schools) is often referred to as multilevel mediation analysis and has received much attention recently (MacKinnon, 2008; Preacher, Zyphur, & Zhang, 2010; Tofighi, West, & MacKinnon, 2013; Zhang, Zyphur, & Preacher, 2009). Clustered data may involve individuals who are nested within clusters or repeated measures collected from the same individuals over time.\(^1\) Data collected at the lower level (e.g., individual) are called Level 1 data, while data collected at a higher level (e.g., cluster) are called Level 2 data. Multilevel data may involve more than two levels. Multilevel mediation techniques model mediation at different levels of the analysis while taking into account bias in standard errors resulting from a lack of independence among observations common in such data (Kenny, Korchmaros, & Bolger, 2003; Krull & MacKinnon, 1999, 2001).

The purpose of this article is to provide a gentle introduction to single-level and multilevel mediation analysis. We will use two numerical examples to demonstrate how to specify, estimate, and interpret results of mediation analysis to answer important research questions in early adolescence research. We will provide accompanying computer code for the Mplus software (L. K. Muthén & Muthén, 1998-2012). The two numerical examples use simulated data taken from the results of published studies.

### Single-Level Mediation Model

This section provides an introduction to the basic single-mediator model with single-level data. The basic single-mediator model is used as a pedagogical tool to describe fundamental concepts in a mediation model that include calculation of direct and indirect effects. Next, we describe a more general case of a mediation analysis with multiple mediators and multiple outcomes in a structural equation modeling (SEM) framework. The section concludes with
a description of a numerical example simulated from a published study in early adolescence research that involves stating research hypotheses, identifying indirect effects based on the hypotheses, and testing indirect effects and interpreting the results.

**A Basic Single-Mediator Model**

Initially, we consider a single-mediator model in which all the observed variables are continuous. Variable $X$ is the independent variable, $M$ is the mediator, and $Y$ is the outcome (dependent) variable. The path diagram for this model is shown in Figure 1. Equations corresponding to this model are as follows:

$$Y = d_1 + cX + \varepsilon_1,$$  \(1\)

$$M = d_2 + aX + \varepsilon_2,$$  \(2\)

$$Y = d_3 + c'X + bM + \varepsilon_3.$$  \(3\)

Coefficient $c$ denotes the total effect of $X$ on $Y$, coefficient $a$ quantifies the effect of $X$ on $M$, coefficient $c'$ is the direct effect of $X$ on $Y$ (the effect of $X$ on $Y$ independent of $M$).
that is not transmitted through $M$), and coefficient $b$ describes the effect of $M$ on $Y$ holding $X$ constant. Note that Equations 1 to 3 represent the classic formulation of a single-mediator model shown in Figure 1 (Baron & Kenny, 1986; MacKinnon, 2008). The terms $e_1$, $e_2$, and $e_3$ denote the residuals and $d_1$, $d_2$, and $d_3$ represent intercepts.

Referring again to Figure 1, the indirect effect of $X$ on $Y$ is measured by the product of two coefficients, $ab$. The direct effect, $c'$, is the effect of $X$ on $Y$ that is not mediated by $M$. Finally, for mediation models with continuous dependent variables and no interaction between $X$ and $M$, the total effect is equal to the sum of the indirect and the direct effects, $c = ab + c'$ (MacKinnon, Warsi, & Dwyer, 1995). For such models, an alternative method to calculating the indirect effect is the difference in coefficients $c - c'$, which is equal to $ab$ (MacKinnon et al., 1995).

Causal Interpretation of Direct and Indirect Effects

So far, we have discussed mediation models and the term “effect” in a mostly statistical sense. An important application of mediation analysis is to examine the mechanism by which an intervention achieves a desired outcome by targeting a mediator in a randomized trial. Imagine that the single mediator model in Figure 1 is based on a randomized trial, where $X$ denotes random assignment to intervention and control group. In a randomized experiment, the intervention effect on the mediator, $a$, and the total intervention effect on the dependent variable, $c$, can be interpreted causally due to randomization of $X$ values. However, coefficients $b$ (path from $M$ to $Y$) and $c'$ (direct effect) in general do not have a causal interpretation because the participants were not randomly assigned to different values of the mediator $M$ (Holland, 1986). When the values of the mediator are not randomized, one cannot rule out the biasing effect of an omitted covariate, for example, some pretest measures that have effects on $M$ and $Y$. The presence of such confounding variables will induce biases in the estimate of $b$ and $c'$. In addition, when $X$ itself is also not randomized, the effect $a$ and the total effect $c$ are usually prone to bias due to omitted covariates. In a recent paper, Green, Ha, and Bullock (2010) criticized that many researchers performing mediation analysis tend to ignore the potentially biasing effects of omitted covariates and implicitly assume unconfoundedness of paths in the mediation model. This point has been previously raised by other researchers, for example, Judd and Kenny (1981) or MacKinnon (2008, chap. 13). The unconfoundedness condition has been formalized by Imai, Keele, and Tingley (2010) under the name of sequential ignorability. Pearl (2011, 2012) has provided additional insights into the
analysis of causal effects in mediation models and B. O. Muthén (2011) has provided Mplus computer code to estimate effects described by Pearl (2011).

**Categorical Mediators**

The purpose of this section is to provide a brief introduction along with references to mediation analysis when the mediator is binary. Clearly, there are situations in which one (or even both) of M and Y might be binary (or categorical with more than two categories). A widely used approach is logistic (or probit) regression for the portions of the model that have binary dependent variables (Hosmer, Lemeshow, & Sturdivant, 2013). For example, when the mediator is binary (M = 0,1), a logistic regression equation for the mediator is specified as follows:

$$\log\left(\frac{p}{1-p}\right) = d_2 + aX,$$

(4)

where $p$ is the probability of the mediator $M$ being equal to 1; this probability is sometimes referred to as the probability of being a “case”; $d_2$ is the intercept and $a$ is the logistic regression coefficient. The ratio $p/(1-p)$ is called the odds of being a case, and the natural logarithm of the odds is usually referred to as the **logit**. Using the logit metric for an outcome variable has the advantage that the right-hand side of the regression equation becomes linear. In Equation 4, $d_2$ and $a$ are in the logit metric. If $a$ is positive (negative), then the logit of being a case and thus the probability of being a case increases (decreases) when the predictor $X$ increases (decreases). If the coefficient $a$ is zero, there is no relationship between $X$ and $M$.

When a mediator is categorical and the outcome variable is continuous, traditionally the indirect effect is calculated in two ways (MacKinnon, 2008, chap. 11): (a) using a product term $ab$ or (b) using a difference $c-c'$ where $a$ is in the logit metric (see Equation 4) and $b$, $c'$ and $c$ are in the metric of the original $Y$ variable (see Equations 1 and 3). Because the two methods of computing an indirect effect use coefficients with different metrics, this poses a computational problem. More specifically, the two methods of computing indirect effects yield different results (MacKinnon & Dwyer, 1993). MacKinnon and Dwyer recommended using the product term $ab$ method and generally discourage the use of the difference in coefficient estimator. However, MacKinnon and Dwyer also showed that when standardized coefficients are used in the computation, the difference between the two methods is usually small and estimates of the indirect effect across the two methods are often very close to each other.
More recently, attempts have been made to more formally express indirect effects in the presence of categorical mediators and/or outcomes. Much of this work is based on defining causal direct and indirect effects using the so-called potential outcomes framework (Rubin, 2005) or graphical models (e.g., Pearl, 2012). One of the most comprehensive treatments of categorical mediators and/or outcomes is given by Valeri and VanderWeele (2013), who derive formulaic expressions for the estimation of direct and indirect effects. In addition, Valeri and VanderWeele also provide SPSS and Statistical Analysis System (SAS) macros that allow applied researchers to estimate direct and indirect effects whenever categorical variables are involved in a mediation model. One also can use the R package mediation (Tingley, Yamamoto, Keele, & Imai, 2013) or an Mplus program as detailed by B. O. Muthén (2011) to conduct causal mediation analysis with categorical variables. These software solutions offer almost identical capabilities in estimating indirect effects for mediation models with categorical variables. Valeri and VanderWeele (2013) provide a detailed comparison of the functionalities of all available software programs.

**Mediation Analysis With Multiple Mediators and Outcomes**

A general approach to test mediation models in single level data is to use SEM (Judd & Kenny, 1981). Using SEM, a researcher can extend the single mediator model described previously to simultaneously specify and estimate relationships between multiple mediators, outcome variables, moderators, and covariates. In an intervention mediation program, it is often the case that the program targets multiple protective and risk factors that, in turn, are expected to affect multiple behavioral outcomes. In addition, researchers often control for various covariates (e.g., pretest measures of mediators or outcomes). Moreover, researchers may be interested in examining the effect of moderators on the indirect effect by including background variables of interest (e.g., gender).

Another advantage of using SEM is the ability to model the measurement error in mediator and outcome variables. Estimates of indirect and direct effects will be attenuated if the measurement errors in mediator and outcome variables are not taken into account. A common approach to account for measurement errors is to construct a latent variable. One can specify a latent variable from multiple items (indicator variables) for relevant variables in a mediation model in the measurement part of an SEM. In the structural part of the SEM, the relationships between different (latent or observed) variables involved in a mediation model are specified. These coefficient estimates are
subsequently used to calculate direct and indirect effects associated with each mediator in the model.

A Numerical Example

We next apply a single level mediation model to an existing data set. Loukas et al. (2006) examined whether the effect of positive perceived school climate on reducing emotional and behavioral problems (i.e., conduct problems and depressive symptoms) was mediated by students’ sense of connection to their school (school connectedness). School connectedness, a measure of students’ social bindings, was considered to be a plausible mediator between students’ perceived school climate and early conduct problems and depressive symptoms among students. The final sample consisted of 489 students, aged 10 to 14, who attended sixth and seventh grades from three middle schools and participated in a two-wave data collection. The students answered a questionnaire with 162 items at Wave 1 and 161 items at Wave 2 in a supervised session.

The students’ perceived school climate construct was considered to have four dimensions: friction, cohesion, competition among students, and overall satisfaction with classes. To measure the perceived school climate at Wave 1, four subscales of a modified version of My Class Inventory (Anderson, 1973) were used. Each subscale had five items and the response to each item was on a 5-point scale with 1 indicating “not at all true” and 5 “very true.” To measure the school connectedness construct, students responded to five questions asking whether they felt closeness and a sense of belonging to their school at Wave 1. The response was on a 5-point scale with 1 indicating “strongly disagree” and 5 indicating “strongly agree” and higher scores showed higher levels of school connectedness. Two aspects of emotional and behavioral problems were measured at Waves 1 and 2: conduct problems and depressive symptoms. A 5-item subscale of the Strength and Difficulties Questionnaire (Goodman, Meltzer, & Bailey, 1998) measured the conduct problems. The answers were on a 3-point scale ranging from 1 indicating “not true” to 3 indicating “certainly true.” Twenty-six items from the Children’s Depressive Inventory (Fraser, 1982) measured the depressive symptoms. The answers were on a 3-point scale ranging from 0 to 2. For each item, a student chose a response among three responses that best described her or his situation. Students’ scores on each scale (subscale) were averaged to show a higher level of the respective construct. For instance, the scores on a 5-item subscale measuring friction were averaged so that higher scores showed higher levels of friction among students.
Research Questions

For this analysis, the research question was whether the relationship between students’ perceived school climate and behavioral and emotional symptoms was mediated by the school connectedness. The independent variables were the four measures of the perceived school climate at Wave 1: friction, cohesion, competition among students, and satisfaction with classes. In addition, Loukas et al. controlled for measures of depressive symptoms and conduct problems at Wave 1. The mediator was the students’ school connectedness measured at Wave 1 and the outcome (dependent) variables were depressive symptoms and conduct behavior problems measured at Wave 2. We quickly add that other potentially confounding variables may exist that were not assessed and included in the model. Even though pretests are important covariates that tend to reduce bias substantially (Steiner, Cook, Shadish, & Clark, 2010), residual confounding due to omitted variables may still occur. The fact that we have longitudinal data helps to rule out certain alternative causal explanations (e.g., the outcome at Wave 2 influencing the mediator at Wave 1), but does not mitigate potential bias due to unobserved confounding variables.

When data on individual items are available, it is preferable to use latent variables instead of average (composite) scores for each construct to account for measurement errors. However, for this example, we only had access to the summary descriptive data in the form of means, standard deviations, and correlations between variables. We used the descriptive statistics to simulate data for each construct composite score.

Model Estimation

We specified an SEM to test hypotheses about the indirect effects stated in the previous section. SEM provides a suitable framework to estimate model parameters required to test indirect effects especially when the model includes multiple mediators and outcomes. Any SEM software package may be used to test the model in Figure 1. We used Mplus (L. K. Muthén & Muthén, 1998-2012) for this analysis.

We should be cognizant of a few important characteristics of the data when conducting a mediation analysis. One characteristic is the scale of the mediator and dependent variables in the model. For this model, mediator and outcome variables are composite scores and therefore are considered to be continuous. A second characteristic is whether the sample has missing data. Depending on the underlying assumptions of missing data mechanisms, one may have to choose a different modeling strategy (Baraldi & Enders, 2010).
If the missing data mechanism is missing at random (MAR; sources of missing data are recorded and are included in the model) or missing completely at random (MCAR; causes of missing data are completely random), one can use the full information maximum likelihood (FIML) estimator (Graham, 2003). For this data set, we do not have missing data and the mediator and outcome variables are continuous. We therefore used the maximum likelihood (ML) estimation. The results of the analysis for only significant path coefficients are shown in Figure 2.

Figure 2. Unstandardized coefficients and standard errors in parentheses for the single-level mediation analysis of the school climate example. Nonsignificant paths are not shown. Correlations between the independent variables and correlation between the dependent variables are not shown. W1 = Wave 1.

*p ≤ .05  **p ≤ .01.
Testing Indirect Effects

A recommended method to test the hypothesis about a specific indirect effect, $H_0: ab = 0$, is to calculate a confidence interval (CI) for an indirect effect as the CI gives a range of plausible values for the indirect effect (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). The CI also has an easy relationship to classic null hypothesis testing. If the CI does not include zero, we may conclude that the indirect effect is significantly different from zero at a given significance level. In this section, we discuss a number of popular methods to test an indirect effect and make a recommendation on which method to use based on the results of past simulation studies.

A popular approach to test the null hypothesis $H_0: ab = 0$ is to compute an asymptotic normal theory CI that is based on the asymptotic properties of the ML estimator (Sobel, 1982). The asymptotic normal theory CI uses the estimate of the indirect effect, $\hat{ab}$, its standard error, $SE(\hat{ab})$, and standard normal distribution critical values to build a symmetric CI. For example, a 95% CI is as follows: $\hat{ab} \pm 1.96SE(\hat{ab})$. A general method to calculate the standard error of an indirect effect is to use the multivariate delta method (Sobel, 1982). Another approach is to use the analytical formula to calculate the standard error as proposed by Craig (1936). The former method is implemented in Mplus while the latter is available in the RMediation package (Tofighi & MacKinnon, 2011).

The main problem with the asymptotic normal theory CI is that the product of two normally distributed random variables is in most circumstances not normally distributed, but, in fact, has a skewed distribution (Craig, 1936). Asymptotic normal theory CIs are likely to produce inaccurate results unless the sample size is sufficiently large (MacKinnon et al., 2002). We do not recommend using this method for single-level analysis. We showed the results in Table 1 for pedagogical purposes.

Another popular method to calculate a CI for an indirect effect is to use general resampling techniques known as the bootstrap method. One resampling method is the nonparametric bootstrap in which a number of repeated samples ($B$) are drawn from the data set. The general logic of the bootstrap is as follows: The mediation model and indirect effects of the original sample are computed along with parameter estimates of $B$ bootstrap samples. This results in $B + 1$ estimates for each path coefficient and indirect effect in the model. The $B + 1$ bootstrap samples approximate the sampling distribution of the parameters and the indirect effects in the model. There are several methods to compute CIs of the parameters and indirect effects using the bootstrap samples: (a) percentile bootstrap, (b) bias-corrected bootstrap (BC), and (c) accelerated bias-corrected (BCa) bootstrap (MacKinnon, Lockwood, &
To compute lower and upper confidence limits, the percentile bootstrap method uses the lower and upper quantiles of the bootstrap samples corresponding to the probabilities $\alpha/2$ and $1 - \alpha/2$, respectively. BC and BC$_{a}$ adjust for potential bias in the percentile CIs (Efron, 1987). For an indirect effect, simulation studies have shown that the percentile bootstrap CIs yield more accurate Type I error rates while BC and BC$_{a}$ CIs resulted in an inflated Type I error rate for a number of combinations of parameters involved in an indirect effect (Biesanz, Falk, & Savalei, 2010; Fritz, Taylor, & MacKinnon, 2012). For this example, we calculated percentile bootstrap CIs for the indirect effects in our model as shown in Table 1.

A better approach to build a CI for an indirect effect is to analytically estimate quantiles of the distribution of the product of two normal random variables. This method is referred to as the distribution of the product of the coefficients method (MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon et al., 2002). The distribution of the product of the coefficients CIs have been shown to be a more powerful test of an indirect effect compared with the asymptotic normal theory and bootstrap CIs (MacKinnon et al., 2002; Tofighi & MacKinnon, 2011). One can use the RMediation package in R (R Development Core Team, 2011) or the web application for the RMediation package, available at http://www.amp.gatech.edu/RMediation, to obtain a distribution of the product of the coefficients CI (Tofighi &

### Table 1. 95% CIs for Indirect Effects for the School Climate Data ($n = 847$).

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Distribution of product</th>
<th>Asymptotic$^a$</th>
<th>Bootstrap$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$LL$</td>
<td>$UL$</td>
<td>$LL$</td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td>Satisfaction</td>
<td>-0.0185</td>
<td>-0.0002</td>
<td>-0.0178</td>
</tr>
<tr>
<td></td>
<td>Friction</td>
<td>0.0001</td>
<td>0.0117</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>Cohesion</td>
<td>-0.0141</td>
<td>-0.0001</td>
<td>-0.0135</td>
</tr>
<tr>
<td></td>
<td>Competition</td>
<td>-0.0039</td>
<td>0.0023</td>
<td>-0.0034</td>
</tr>
<tr>
<td>Conduct problems</td>
<td>Satisfaction</td>
<td>-0.0283</td>
<td>-0.0041</td>
<td>-0.0274</td>
</tr>
<tr>
<td></td>
<td>Friction</td>
<td>0.0021</td>
<td>0.0181</td>
<td>-0.0008</td>
</tr>
<tr>
<td></td>
<td>Cohesion</td>
<td>-0.0217</td>
<td>-0.0031</td>
<td>-0.0208</td>
</tr>
<tr>
<td></td>
<td>Competition</td>
<td>-0.0062</td>
<td>0.0038</td>
<td>-0.0057</td>
</tr>
</tbody>
</table>

Note. Distribution of the product = Distribution of the Product of Coefficients; Asymptotic = Asymptotic Normal Theory CI; Bootstrap = Percentile Bootstrap; Satisfaction = Satisfaction With Classes; Friction = Friction Among Students; Cohesion = Cohesion Among Students; Competition = Competition Among Students; $LL$ = Lower Limit of CI; $UL$ = Upper Limit of CI. CI = confidence interval.

$^a$CIs are produced by RMediation that uses SEs that are slightly different from the multivariate delta SEs produced by Mplus.

$^b$CIs are produced by Mplus with three decimal points.
MacKinnon, 2011). To compute a distribution of the product of the coefficients CI using the RMediation web application, one needs to obtain coefficient estimates and their standard errors from an SEM analysis. For this example, we used the coefficient estimates and their standard errors from the Mplus analysis output. The significant paths are shown in Figure 2. To obtain the distribution of the product of the coefficient CIs, correlations between coefficient estimates of the paths that constitute the indirect effects were assumed to be zero.

For pedagogical purposes, we reported the results of the distribution of the product of the coefficients, percentile bootstrap, and asymptotic normal theory 95% CIs to test an indirect effect in Table 1. For the final results, however, we recommend reporting either the distribution of the product of the coefficients or percentile bootstrap CIs and interpret the results accordingly. We do not recommend using the asymptotic normal theory method when the alternative approaches are available.

**Results**

As can be seen in Table 1, the results from the percentile bootstrap and distribution of the product of the coefficient methods agree with each other for the most part. Based on the distribution of the product of coefficients CIs (or bootstrap) method, the indirect effects of satisfaction with classes, friction among students, and cohesion among students on depressive symptoms at Wave 2 were significantly different from zero at the .05 significance level. In addition, the indirect effects of satisfaction with classes, friction among students, and cohesion among students on conduct problems at Wave 2 were significantly different from zero at the .05 significance level, based on the distribution of the product of coefficients CIs (or bootstrap) method.

For comparison purposes, we discuss the asymptotic normal theory results. Based on the asymptotic normal theory CIs, none of the indirect effects of satisfaction with classes, friction among students, and cohesion among students on depressive symptoms at Wave 2 were significantly different from zero. Only the indirect effect of satisfaction with classes on the conduct problems was significantly different from zero. We therefore recommend that researchers report the final results based on the distribution of the product of coefficients or bootstrap method.

Consequently, based on the distribution of the product of coefficients CIs (or bootstrap) method, it appears that school connectedness did mediate the effects of the three dimensions of the perceived school climate (satisfaction with classes, friction among students, and cohesion among students) on
Tofighi and Thoemmes

Depressive symptoms and conduct problems at Wave 2. The indirect effects of satisfaction with classes and cohesion among students were negative on conduct problems and depressive symptoms. This indicates that as satisfaction with classes and cohesion among students increase, students’ connection to school increases, which in turn would lead to students having a decreased level of depressive symptoms and fewer instances of conduct problems. Indirect effects of friction among students on the dependent variables were positive indicating that as friction among students increases, conduct problems and depressive symptoms would also increase. Finally, competition among students did not have a significant indirect effect through school connectedness on conduct problems and depressive symptoms.

Multilevel Mediation Analysis

In adolescence research, many research studies occur in cluster settings such as schools and classrooms, where a researcher is interested in examining indirect effects at the cluster level, individual level, or both (Espelage, Holt, & Henkel, 2003; Patrick, Ryan, & Kaplan, 2007; Torsheim & Wold, 2001; Vieno et al., 2007). For example, in school settings, data collection may occur at two levels that involve students (Level 1) and classrooms (Level 2). A researcher might be interested in examining the effect of an environmental factor (e.g., instructor attitude toward bullying measured at Level 2) on a mediator (e.g., student’s intention to stop bullying measured at Level 1) that, in turn, affects student’s behavior (e.g., victimization measured at Level 1; Hektner & Swenson, 2012). This is called a $2 \rightarrow 1 \rightarrow 1$ model in which the first, second, and third numbers correspond to measurement levels of the independent, mediator, and outcome variable, respectively. In a multilevel mediation study, data on independent, mediator, and outcome variables may be measured at student (individual) level, classroom (cluster) level, or both. Other types of mediation models are also possible, such as a $1 \rightarrow 1 \rightarrow 1$ model, in which the variables in the model are all measured at Level 1. Furthermore, a multilevel mediation study may involve data with more than two levels; for instance, consider a cluster setting with three levels of nesting in which students are nested within classrooms, and classrooms are nested within schools. The focus of the current study is on a two-level $2 \rightarrow 1 \rightarrow 1$ mediation model as it is common in cluster-randomized trials as well as studying adolescent behaviors in school settings where the focus of research is on the processes that occur at the cluster level (e.g., classroom).

One important decision in examining indirect effects in a multilevel mediation model is choosing the type of centering for mediators measured at Level 1. In a two-level $2 \rightarrow 1 \rightarrow 1$ model, a researcher has two options to center the
mediator that would lead to two separate interpretations: (a) CWC (centering within cluster) 2 and (b) CGM (centering at the grand mean) 2 (Enders & Tofghi, 2007; Kreft, de Leeuw, & Aiken, 1995; Pituch & Stapleton, 2012). CWC2 centers the mediator within clusters (CWC) using cluster means. These cluster means are then added as predictors at Level 2. CGM2 centers mediator variables at the grand mean (CGM) while adding the cluster means of the mediator at a Level 2 predictor.

The result of applying CWC2 to a 2→1→1 model, shown in Figure 3, is that CWC2 decomposes the effect of the mediator on the outcome into between- and within-cluster effects. Because the independent variable \( X \) does not vary at the individual level, only between-cluster indirect effects are estimated. CWC2 is appropriate when the individual level mediator variable is used to measure a cluster level construct and thus the focus of research is at the cluster level (Pituch & Stapleton, 2012). An example of this type of research is when individuals within a cluster rate a shared environmental characteristic at the cluster level. For example, consider the numerical example discussed next in which students (Level 1) were asked to rate their teacher’s level of activity in a classroom (Level 2) to stop or prevent bullying. In this example, the aggregate measure of student ratings are used to measure a teacher’s behavior, which is a cluster (classroom) level construct. When
using CWC2 to center a mediator, the indirect effect can only exist at the cluster level (Zhang et al., 2009).

Applying CGM2 to a Level 1 mediator is appropriate when a researcher is interested in examining the effect of the independent variable on the mediator at the individual level and the mediator represents the measurement of an individual’s characteristic on an absolute scale rather than on the person’s relative position within a cluster as measured by his or her deviation score from the cluster mean. For a $2 \rightarrow 1 \rightarrow 1$ model, CGM2 centering decomposes the indirect effect into two parts: a within-cluster indirect effect through the within-cluster part of the mediator and a unique between-cluster indirect effect through the aggregate (cluster means) of the mediator. The unique between-cluster indirect effect represents a contextual indirect effect that is above and beyond the within-cluster indirect effect (Pituch & Stapleton, 2012).

In CWC2, the cluster means can be either observed or latent means (Lüdtke et al., 2008). The choice of using observed or latent cluster means in CWC2 depends on the context; in general, using latent cluster means yields less attenuated coefficient estimates while using observed cluster means provides smaller standard errors of the estimates. A general approach to estimate the multilevel mediation model is to use the multilevel structural equation modeling (MSEM) framework (MacKinnon, 2008; Preacher et al., 2010).

### A Numerical Example

This section describes how to conduct a multilevel mediation analysis. For this numerical example we use simulated data generated based on an existing study by Hektner and Swenson (2012). Hektner and Swenson examined an indirect effect of teachers’ beliefs about bullying (independent variable) on students’ actual victimization (outcome variable) in schools through teachers’ intervention to stop bullying (mediator). The simulated data contained a sample of 400 eighth-grade students (Level 1 units), ages from 8 to 15, selected from 40 classes (Level 2 units) with their respective teachers. For the simulated data set to have a strict hierarchical structure, classes did not share a teacher unlike the actual data set where some teachers taught more than one class. Therefore, data for 40 teachers and 40 classes were generated.

Students answered eight questions about victimization with each question having five response options measuring the frequency with which they have been bullied in the past couple of months, with 1 indicating no bullying and 5 indicating being bullied several times a week. Teacher intervention to stop bullying was a composite measure of two 5-point items rated by students with a higher score indicating a higher level of activity by a teacher in the
classroom to stop or prevent bullying. To measure beliefs about bullying, teachers answered a questionnaire that measured their beliefs that bullying was a normative developmental part of being an adolescent. We simulated composite scores for normative beliefs, with a higher score indicating a higher level of the normative beliefs.

**Research Question**

The research question was whether the effect of teacher beliefs about bullying (independent variable) on student victimization (outcome variable) was mediated by teacher intervention (mediator). Because teacher beliefs were measured at Level 2, this was a $2 \rightarrow 1 \rightarrow 1$ mediation model. It was hypothesized that teacher beliefs about bullying led to teacher intervention, which led to student victimization. The more a teacher considers bullying as normative, the less likely the students considered the teacher as intervening to stop bullying. In addition, lower levels of teacher intervention were hypothesized to predict higher levels of student victimization. As in the previous example, effects may be biased due to omitted confounders at either level of the analysis.

**Model Estimation**

The first step to consider when conducting a multilevel analysis is to test whether there is a significant clustering effect and thus the use of multilevel modeling is needed. This is determined by testing whether intraclass correlation coefficients (ICCs) for Level 1 variables equal zero. ICCs indicate the degree to which scores from the same cluster are correlated with each other. ICCs range from zero (no correlation) to one (perfect correlation between observations). If ICCs are greater than zero, a multilevel model would produce more accurate results (e.g., $SE$ and $p$ values from significance tests) than ordinary least squares (OLS) regression that ignores the clustering. The next step is to decide the scaling of the mediator variables measured at a lower level (e.g., Level 1; Enders & Tofighi, 2007; Kreft et al., 1995; MacKinnon, 2008; Preacher et al., 2010). As mentioned previously, the choice of centering for the mediator depends on the research question and whether the mediator represents a classroom (Level 2) versus a student-level construct (Pituch & Stapleton, 2012). In this example, we used the CWC2 with latent cluster means centering strategy because the students rated the mediating variable of teachers’ level of activities in classrooms to stop or prevent bullying, and therefore the teacher intervention construct was deemed a classroom level construct. In the CWC2 with latent cluster means centering strategy, latent
cluster means were also added as Level 2 predictors (L. K. Muthén & Muthén, 1998-2012, pp. 262-263).

As shown in Figure 3, applying CWC2 with latent cluster means to observed variables measured at Level 1 (i.e., $M_{ij}$ and $Y_{ij}$) creates latent constructs at the between- and within-cluster levels shown by circles. For example, the observed mediator, teacher intervention ($M_{ij}$), was decomposed into two orthogonal latent mediators: (a) a between-cluster latent mediator, $\eta_{Mj}$, that represents a teacher intervention construct at the classroom level, and (b) a within-cluster latent variable, $\eta_{Mij}$, that represents the relative standing of a particular student’s rating of a teacher relative to the other students’ average rating in a classroom. It can be shown that $M_{ij} = \eta_{Mj} + \eta_{Mij}$ (L. K. Muthén & Muthén, 1998-2012).

In addition, a researcher needs to decide if a relationship at the within-classroom part of the model (i.e., the within-classroom effect of teacher intervention on student victimization denoted by $b_w$) is to be fixed or allowed to vary randomly. For our example, we used a random intercepts multilevel model, that is, $b_w$ is a fixed regression slope. This is done for two reasons: (a) because in a $2\rightarrow1\rightarrow1$ model, we were interested in the between-classroom effect, and (b) we chose a more parsimonious random intercepts model compared with a random slopes model that requires a more comprehensive Level 2 covariance structure. The random intercepts model specifies the between-cluster regression equations while within-cluster regression coefficients (slopes) are considered fixed. Fixed slopes refer to the fact that within-classroom regression coefficients are assumed to be constant across classrooms. In words, the effect of teacher intervention (mediator) on student victimization (outcome variable) was not allowed to vary across classrooms because we were mainly interested in the between-classroom effect of teacher intervention on student victimization controlling for teacher beliefs about bullying.

One conventional way to write a multilevel mediation model is to use Level 1 (student level) and Level 2 (classroom) equations. As noted before, the MSEM framework in Mplus decomposes the Level 1 variables, $M_{ij}$ and $Y_{ij}$, into orthogonal between- and within-classroom parts. Subscripts $i$ and $j$ denote student and classroom, respectively. Below are the Level 1 and Level 2 equations for the mediator teacher intervention ($M_{ij}$):

Level 1:

$$M_{ij} = \eta_{Mj} + \eta_{Mij}.$$  \hspace{1cm} (5)

Level 2:

$$\eta_{Mj} = \mu_{Mj} + a_b X_j + \zeta_{Mj}.$$  \hspace{1cm} (6)
In the Level 1 equation, students’ rating of teacher intervention equals the random intercept, $\eta_{Mj}$, plus a student level random error, $\eta_{Mij}$, where $\eta_{Mj}$ is the true (latent) classroom $j$’s mean rating of teacher intervention that represents a teacher intervention construct at the classroom level. $\eta_{Mij}$ represents student $i$’s deviation score from her classroom mean; it represents a standing of a particular student’s rating of a teacher relative to the other students’ average rating in a classroom. In the MSEM framework, $\eta_{Mj}$ and $\eta_{Mij}$ are latent variables that capture between- and within-classroom variation, respectively.

In fact, Equation 5 is a familiar random effects ANOVA that estimates the between- and within-classroom variance components. Within-classroom (Level 1) residual variance, $\text{Var}(\eta_{M}) = \theta_{M}$, captures the variation of each student’ score from her true classroom mean score. The between-classroom variance, $\text{Var}(\eta_{Mj}) = \tau_{0M}$, quantifies variation in the classroom mean scores. We can use the results of the random effects ANOVA to estimate the variance of the between- and within-classroom parts of the mediator and then compute an ICC as follows: $\text{ICC} = \tau_{0M} / (\tau_{0M} + \theta_{M})$. In this formulation, the ICC estimates the proportion of the total variance of the mediator that is at the classroom level. If the ICC is zero, it implies that the mediator varies entirely at Level 1 and thus use of a multilevel model for the mediator is unnecessary. If the ICC is greater than zero, then the use of multilevel modeling is justified. One can test if the ICC is zero by testing whether the between-classroom variance, $\tau_{0M}$, is zero, $H_{0} : \tau_{0M} = 0$ (Snijders & Bosker, 2011, chap. 3). To test this hypothesis, we used the $F$ test statistic from the random effects ANOVA using SPSS.

In Equation 6, the true classroom mean rating of teacher intervention is a function of the overall mean, $\mu_{Mj}$, the effect of classroom mean rating of teacher beliefs on teacher intervention, $a_{b}$, and the random deviation of true classroom mean score from the predicted score, $\zeta_{Mj}$. Level 2 residual variance, $\text{Var}(\zeta_{Mj}) = \psi_{M}$, captures between-classroom residual variance of $\eta_{Mj}$ after controlling for the effect of $X_{j}$.

Next, we specified the multilevel equations for the outcome variable student victimization as follows:

Level 1:

\[
Y_{ij} = \eta_{Yj} + \eta_{Yij}, \quad (7)
\]

\[
\eta_{Yj} = b_{w} \eta_{Mij} + \zeta_{Yj}. \quad (8)
\]

Level 2:

\[
\eta_{Yj} = \mu_{Yj} + c'_{b} X_{j} + b_{b} \eta_{Mj} + \zeta_{Yj}. \quad (9)
\]
In Equation 7, a student’s victimization score is decomposed into two orthogonal latent variables, \( \eta_{Yj} \) and \( \eta_{Yij} \) that capture between- and within-classroom variation, respectively. The random intercept \( \eta_{Yj} \) represents the true mean of the victimization score for classroom \( j \). The random error term \( \eta_{Yij} \) quantifies a student \( i \)'s deviation score from her classroom true mean score. The Level 1 residual variance, \( \text{Var}(\eta_{Yij}) = \theta_Y \), quantifies the variation of each student’s score from her true classroom mean score. The between-classroom variance, \( \text{Var}(\eta_{Yj}) = \tau_{0Y} \), quantifies variation in the classroom true mean scores. We used the variance of the between- and within-classroom parts of the outcome variable to estimate the ICC.

Equation 8 shows the within-class relationship between the outcome variable and the mediator. Coefficient \( b_w \) denotes the within-classroom effect of teacher intervention on victimization. It should be noted that \( b_w \) does not vary across clusters in Equation 8; the within-classroom coefficient is assumed be constant across classrooms. This particular effect was fixed because we chose to use a more parsimonious model. \( \zeta_{Yij} \) is a Level 1 residual term.

However, if a researcher has a theoretical reason to estimate a random slope (a slope that varies across classrooms), she would need to change the fixed slope to a random slope, \( b_{wj} \), in Equation 8. Then, she needs to introduce a new equation for the random slope as follows:

\[
b_{wj} = b_w + \zeta_{b_{wj}},
\]

where \( \zeta_{b_{wj}} \) captures deviation of the random slope, \( b_{wj} \), from the average slope, \( b_w \).

In Equation 9, \( \mu_{Yj} \) is the intercept, \( b_b \) denotes the between-classroom effect of teacher intervention on student victimization controlling for teacher beliefs about bullying, \( c_b^b \) quantifies the between-classroom direct effect of teacher beliefs on student victimization, the random error term \( \zeta_{Yj} \) is the between-classroom residual for the classroom victimization score. Finally, the Level 2 residual variance, \( \text{Var}(\zeta_{Yj}) = \psi_Y \), captures between-classroom residual variance of \( \eta_{Yj} \) after controlling for the effects of \( X_j \) and \( \eta_{Mj} \).

Next, we estimated the between-classroom indirect effect of teacher beliefs about bullying on student victimization using the product of two coefficients: \( a_p b_b \). The within-classroom effect of teacher beliefs about bullying (\( X_j \)) on teacher intervention (\( M_{ij} \)) could not be estimated because teacher beliefs about bullying was a Level 2 variable that did not vary at the student level (Preacher et al., 2010). This does not mean that teacher beliefs did not influence students’ rating of teacher intervention. What it means is that because all the students in a classroom rated the same teacher, the differential
effect of teacher beliefs about bullying on students’ rating of the teacher intervention could not be estimated at the student level. We conducted the analysis using Mplus that accommodates the CWC2 with latent cluster mean centering strategy within the general framework of MSEM. We used the maximum likelihood robust (MLR) estimator in Mplus, which is the ML estimator with standard errors that are robust to nonnormality. The robust standard errors are computed using a sandwich estimator (White, 1980). The results are shown in Table 2.

### Testing Indirect Effects

In classical statistics, two available methods to compute CIs for an indirect effect in a multilevel mediation model are the asymptotic normal theory and the distribution of product of the coefficients. Bootstrap (resampling) methods are not yet available to test an indirect effect for a multilevel mediation model. Again, we recommend using the distribution of the product of the coefficients method because it produces more accurate results for smaller sample sizes. Mplus produces asymptotic normal theory CIs. We used the RMediation package to calculate the distribution of the product of the coefficients CIs (Tofighi & MacKinnon, 2011). Similar to the analysis performed in the single-level mediation example, to calculate a distribution of

#### Table 2. Unstandardized Results for the Multilevel Mediation Analysis of Teacher Beliefs About Bullying Example With 40 Teachers (n = 340).

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Parameter</th>
<th>Estimate (SE)</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>Teacher intervention</td>
<td>$a_b$</td>
<td>-0.314 (0.097)</td>
<td>-0.504 -0.124 .001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher beliefs</td>
<td>$b_b$</td>
<td>-0.551 (0.219)</td>
<td>-0.98 -0.122 .012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Victimization</td>
<td>$c'_b$</td>
<td>-0.063 (0.102)</td>
<td>-0.264 0.137 .537</td>
<td></td>
</tr>
<tr>
<td>Indirect effect (mediator = teacher intervention)$^a$</td>
<td>Victimization Teacher beliefs</td>
<td>$a_bb'_b$</td>
<td>0.173 (0.087)</td>
<td>0.028 0.375</td>
<td></td>
</tr>
<tr>
<td>Residual variance</td>
<td>Level 1</td>
<td>$\theta_M$</td>
<td>2.351 (0.177)</td>
<td>2.005 2.697 &lt;.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\theta_Y$</td>
<td>2.106 (0.447)</td>
<td>1.231 2.981 &lt;.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>$\psi_M$</td>
<td>1.255 (0.282)</td>
<td>0.702 1.809 &lt;.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\psi_Y$</td>
<td>1.002 (0.315)</td>
<td>0.385 1.619 .001</td>
<td></td>
</tr>
</tbody>
</table>

Note. CI= confidence interval; LL = lower limit, UL = upper limit. $^a$The distribution of the product of the coefficients CI was produced by RMediation. p values are not available for this method.
the product of the coefficients CI using RMediation, we obtained coefficient estimates, \( \hat{a}_b \) and \( \hat{b}_b \) and their respective standard errors from an MSEM analysis output produced by Mplus. The correlation between the coefficient estimates was assumed to be zero. The results are shown in Table 2 and Figure 3.

**Results**

The results of ANOVA showed that the ICC for the mediator teacher intervention, \( .42, F(39, 360) = 8.173, p < .001 \), and for the outcome victimization, \( .37, F(39, 360) = 6.746, p < .001 \), were greater than zero. The ICC estimates indicate that the students’ scores on the mediator and outcome variables within the same classroom were highly correlated and a considerable proportion of the total variance exists at the classroom level. As a result, use of multilevel modeling is justified and would lead to more accurate results (e.g., SE and CI for the regression coefficients). In the between-classroom part of the model, teacher beliefs about bullying as a normative behavior negatively predicted teacher intervention, \( \hat{a}_b = -0.314, 95\% \text{CI } [-0.504, -0.124], p = .001 \). In turn, teacher intervention was negatively related to the student victimization, \( \hat{b}_b = -0.551, 95\% \text{CI } [-0.98, -0.122], p = .012 \). As a result, the between-cluster indirect effect of teacher beliefs about bullying on student victimization was \( 0.173 (= -0.314 \times -0.551) \). The distribution of the product of the coefficients 95\% CI for the indirect effect was \([0.028, 0.375]\), which did not contain zero. This indicates that the between-classroom indirect effect of teacher beliefs on student victimization through teacher intervention was significantly different from zero. As a teacher’s belief score about bullying as a normative behavior increases by 1 point, the student victimization score increases indirectly (i.e., through teacher intervention) by 0.173 points. The indirect effect ranged from 0.028 to 0.375 points. In addition, the between-classroom direct effect of teacher beliefs on student victimization was not significant, \( \hat{c}_b = -0.063, 95\% \text{CI } [-0.264, 0.137], p = 0.537 \). This indicates that the between-classroom effect of teacher beliefs on student victimization was completely mediated through teacher intervention.

**Summary and Conclusion**

The application of mediation analysis in various areas of social science, including adolescence research, is still thriving and of great interest to applied researchers. Two common applications of mediation analysis are to examine a mediator to explain a causal relationship between two variables, or to use mediation models to design intervention studies that target risk and protective factors that, in turn, are expected to change the behavioral outcomes.
Techniques to conduct mediation analysis using single-level data (i.e., without a clustered structure) have received extensive attention in the literature. Techniques to conduct mediation analysis in multilevel data with a strict hierarchical structure have also begun to become more common.

In this article, we discussed how to conduct and interpret mediation analysis in single-level and hierarchical data sets to answer important research questions in adolescence research. We discussed application of MSEM to conduct multilevel mediation analysis. MSEM allows researchers in adolescence research to examine the measurement structure of psychological constructs in addition to estimating structural relationships among constructs at between- and within-cluster levels. In addition, latent variables within the MSEM framework allow researchers to model the measurement error and thus obtain more reliable estimates of the important psychological constructs.

In sum, mediation analysis is a popular technique as it addresses important questions about how an independent variable impacts behavioral outcomes by changing one or more intervening variables in a causal process. In adolescence research, mediation analysis can particularly help researchers study causal effects and investigate how and why changes occur in adolescent behavior by identifying and targeting important mediators as shown by the two numerical examples.

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

Funding
The author(s) received the following financial support for the research, authorship, and/or publication of this article. Davood Tofighi received no financial support for the research, authorship, and/or publication of this article. Felix Thoemmes was supported by the Institute of Social Science Small Grants program at Cornell University.

Notes
1. In this manuscript, we focus on hierarchical data structure in which a Level 1 unit is nested within one and only one Level 2 unit (i.e., strict hierarchy)
2. This is an active and growing area of research and its complete treatment is beyond the scope of the current manuscript.
3. For a more extensive treatment of this topic and detailed computational instruction on calculating standardized coefficients, please see MacKinnon (2008, chap. 11).
4. The data file (“LoukasData2006.txt”) and Mplus input and output files (“Mplus Analysis.pdf”) are included in the supplemental online materials.
5. For a more thorough discussion of different methods of calculating the SE of an indirect effect, we refer interested readers to MacKinnon, Lockwood, Hoffman, West, and Sheets (2002).
6. Mplus reported results up to three decimal points.
7. For pedagogical purposes, the simulated data have different multilevel structure and sample size from the original data used by Hektner and Swenson (2012).
8. If multiple teachers taught the same classrooms, then the data would have a cross-classified multilevel structure (Raudenbush & Bryk, 2002, chap. 12).
9. If one were to use Mplus to estimate a random effects ANOVA, the estimates of $\theta_M$ and $\tau_0M$ could be slightly different from the ones produced by SPSS because of the different estimation methods used in the software packages.
10. The data file (“Hektner.txt”) and Mplus input and output files (“Mplus Analysis.pdf”) are included in the supplemental online materials.
11. We also recommend a Bayesian approach to estimate an indirect effect in multilevel structural equation modeling (MSEM; Yuan & MacKinnon, 2009). For Mplus implementation of the Bayesian approach, see B. O. Muthén and Asparouhov (2012).
12. The confidence interval (CI) was obtained by plugging in relevant results from the Mplus output file into the RMediation package.

References


*Psychological Methods, 13*, 203-229. doi:10.1037/a0012869


**Author Biographies**

**Davood Tofighi** is currently an assistant professor in the School of Psychology at the Georgia Institute of Technology. Dr. Tofighi’s research interests include the development and application of statistical methodologies and software packages in mediation analysis, Bayesian analysis, and multilevel and longitudinal data analysis applied in prevention research.

**Felix Thoemmes** is currently an assistant professor in the Department of Human Development with joint appointment in the Department of Psychology at Cornell University. His research interests are causal inference, propensity score methods, missing data, and mediation models.