Group Processes & Intergroup Relations

G P I R

Article

Group Processes & Intergroup Relations
2015, Vol. 18(3) 274–289

DOI: 10.1177/1368430214550343
gpir.sagepub.com
SAGE

Investigating multilevel mediation with fully or partially nested data

Mark J. Lachowicz,[1] Sonya K. Sterba,[1] and Kristopher J. Preacher[1]

Abstract

Mediation analysis has become one of the most widely used tools for investigating the mechanisms through which variables influence each other. When conducting mediation analysis with fully nested data (e.g., individuals working in teams) or partially nested data (e.g., individuals working alone in one study arm but working in teams in another arm) special considerations arise. In this article we (a) review traditional approaches for analyzing mediation in nested data, (b) describe multilevel structural equation modeling (MSEM) as a versatile technique for assessing mediation in fully nested data, and (c) explain how MSEM can be adapted for assessing mediation in partially nested data (MSEM-PN) and introduce two new MSEM-PN specifications. MSEM-PN affords options for testing equality of level-specific mediation effects in the nested arm with mediation effects in the nonnested arm. We demonstrate the application of MSEM and MSEM-PN in simulated examples from the group processes literature involving fully and partially nested data. Finally, we conclude by providing software syntax and guidelines for implementation.

Keywords

mediation, multilevel modeling, partial nesting, structural equation modeling

Paper received 5 February 2014; revised version accepted 24 July 2014.

The goal of many studies is to identify and understand the processes through which phenomena occur. This goal is often accomplished by studying the effects of an intervening variable, for example, a variable that transmits the effects of an independent variable to a dependent variable. Investigating the nature of intervening variables, or *mediators*, is commonly known as mediation analysis.

When data are clustered within groups, the independence assumption of ordinary least squares (OLS) regression underlying conventional mediation analysis is violated, leading to potentially biased confidence intervals (CIs). A common example of clustering, or *nesting*, arises in group processes research where individuals are nested within teams (e.g., Nohe, Michaelis,

[1]Vanderbilt University, USA

Corresponding author:
Mark J. Lachowicz, Quantitative Methods Program, Department of Psychology & Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203, USA.
Email: Mark.J.Lachowicz@Vanderbilt.edu

Menges, Zhang, & Sonntag, 2013; van Mierlo, Rutte, Vermunt, Kompier, & Doorewaard, 2007; Zhou, Wang, Chen, & Shi, 2012). Data can sometimes be partially nested as well (e.g., when there is nesting of persons within teams in a treatment arm but no nesting in a control arm; Kirschner, Paas, Kirschner, & Janssen, 2011; Vadasy & Sanders, 2008). This introduces further complexities for assessing mediation.

The purposes of this article are to (a) review traditional approaches for analyzing mediation in clustered data, including single-level and multilevel modeling (b) describe multilevel structural equation modeling (MSEM) as a versatile technique for assessing mediation in fully nested data, (c) describe how MSEM can be adapted for assessing mediation in partially nested data and introduce two new specifications, and (d) demonstrate MSEM in simulated examples from the group processes literature involving fully and partially nested data. Below, when discussing mediation for multilevel designs, we will refer to designs according to the level at which each variable is measured. For example, in a 2-1-2 design the first and last variables in the pathway are measured at Level 2 but the mediator is measured at Level 1.

Mediation Within a Single-Level Modeling Framework

The classic three-variable, single-level mediation model can be viewed as a series of regressions from an independent variable, x_i, to a mediator, m_i, and from m_i to a dependent variable, y_i, controlling for x_i. Often the effects are estimated using the following equations:

$$y_i = \beta_0^y + \beta_1^y x_i + \beta_2^y m_i + r_i^y \quad (1)$$

$$m_i = \beta_0^m + \beta_1^m x_i + r_i^m \quad (2)$$

where: $\begin{bmatrix} r_i^y \\ r_i^m \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_r^{2y} & \\ 0 & \sigma_r^{2m} \end{bmatrix}\right)$

Superscripts for each parameter denote the respective outcome variable. β_0^m and β_0^y are intercepts and r_i^m and r_i^y are normally distributed residuals. The estimate of the mediation effect, or *indirect effect*, is the product of the slope of x_i predicting m_i (i.e., β_1^m) in Equation 2 and the slope of m_i predicting y_i controlling for x_i (i.e., β_1^y) in Equation 1.

Once quantified and estimated, it is usually of interest to gauge significance and precision of the indirect effect $\beta_1^m \beta_1^y$. Many methods have been proposed, but the best-performing ones in terms of power and Type I error rate include using bootstrap CIs (Bollen & Stine, 1990; MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002), Monte Carlo CIs (MacKinnon et al., 2004), Bayesian credible intervals (Yuan & MacKinnon, 2009), and a method of constructing CIs based on the distribution of product terms (MacKinnon, Fritz, Williams, & Lockwood, 2007; Tofighi & MacKinnon, 2011). When data are nested (or partially nested), however, the nonindependence of observations typically leads to downwardly biased estimates of the standard error, overly narrow CIs, increased Type I error rates for the indirect effect (Krull & MacKinnon, 1999), and possibly biased indirect effects (Bauer, Preacher, & Gil, 2006) when single-level methods are employed. These negative consequences of conducting single-level mediation analysis with clustered data are exacerbated as the cluster size and intraclass correlation (ICC) of the mediator and dependent variable (DV) become larger (Krull & MacKinnon, 2001). The problem of nonindependence can be addressed by aggregating data at the group level and proceeding with single-level regression; however, this method is seriously limited by loss of power (Raudenbush & Bryk, 2002). Moreover, researchers using this strategy can fall prey to the ecological fallacy—the use of cluster-level results to make inferences at the individual level.

Mediation Within a Multilevel Modeling Framework

Multilevel modeling (MLM) accommodates clustered data by allowing the simultaneous estimation of coefficients and residual variance at the individual and group levels (respectively, Level 1 and Level 2; Raudenbush & Bryk, 2002). Slopes and intercepts may be allowed to vary randomly across groups. In cross-sectional data, Level 1 units are commonly individuals nested within groups, such as workers within teams. MLM provides a modeling approach for investigating mediation at different levels of the data hierarchy. This includes designs in which x_{ij}, m_{ij}, and y_{ij} are measured at the lowest level (1-1-1 designs; Kenny, Korchmaros, & Bolger, 2003; Krull & MacKinnon, 2001; Pituch, Stapleton, & Kang, 2006; Pituch, Whittaker, & Stapleton, 2005), when x_j is measured at the group level and m_{ij} and y_{ij} are measured at the individual level (2-1-1 designs; Krull & MacKinnon, 1999, 2001; Pituch & Stapleton, 2008; Pituch et al., 2006), and when both x_j and m_j are measured at the group level and y_{ij} is measured at the individual level (2-2-1 designs; Krull & MacKinnon, 2001; Pituch et al., 2006). Here we mainly consider 1-1-1 designs, although much of the discussion to follow also pertains to other designs, such as 2-1-1.

Consider the following Level-1 equations for a multilevel mediation model for a 1-1-1 design:

$$y_{ij} = \beta_{0j}^{y} + \beta_{1j}^{y} x_{ij} + \beta_{2j}^{y} m_{ij} + \varepsilon_{ij}^{y} \tag{3}$$

$$m_{ij} = \beta_{0j}^{m} + \beta_{1j}^{m} x_{ij} + \varepsilon_{ij}^{m} \tag{4}$$

$$\text{where: } \begin{bmatrix} \varepsilon_{ij}^{y} \\ \varepsilon_{ij}^{m} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\varepsilon}^{2y} & \\ 0 & \sigma_{\varepsilon}^{2m} \end{bmatrix} \right)$$

y_{ij} and m_{ij} are the individual-level outcome and mediator for individual i in group j. x_{ij} is the individual-level predictor. β_{0j}^{y} and β_{0j}^{m} are random intercepts allowed to vary across J groups. These random intercepts can be viewed as realizations drawn from normal distributions of possible values. In the Level-2 equations,

$$\begin{aligned} \beta_{0j}^{y} &= \gamma_{00}^{y} + u_{0j}^{y} \\ \beta_{0j}^{m} &= \gamma_{00}^{m} + u_{0j}^{m} \end{aligned} \tag{5}$$

$$\beta_{1j}^{y} = \gamma_{10}^{y} \qquad \beta_{2j}^{y} = \gamma_{20}^{y} \qquad \beta_{1j}^{m} = \gamma_{10}^{m} \tag{6}$$

$$\text{where: } \begin{bmatrix} u_{0j}^{y} \\ u_{0j}^{m} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00}^{y} & \\ 0 & \tau_{00}^{m} \end{bmatrix} \right)$$

γ_{00}^{y} and γ_{00}^{m} are mean intercepts. Slopes of individual-level variables (γ_{10}^{y}, γ_{20}^{y}, γ_{10}^{m}) are fixed effects in Equation 6, but these slopes could be treated as random (not shown here). us are Level-2 error terms. A CI for the indirect effect $\gamma_{20}^{y}\gamma_{10}^{m}$ can be obtained using, for instance, asymmetric Monte Carlo CIs (MacKinnon et al., 2004; Preacher & Selig, 2012). Errors are assumed to be normally distributed with means of zero and are uncorrelated across levels of the data hierarchy.

There has been a progression of developments on how to best fit these models. Initially, such multilevel mediation models were fit as two separate univariate multilevel models, one for y_{ij} and one for m_{ij} (Kenny et al., 2003; Krull & MacKinnon, 2001; Pituch et al., 2006; Pituch et al., 2005). However, under certain circumstances, this univariate approach does not allow estimation of a particular term needed to accurately calculate the indirect effect (i.e., when β_{2j}^{y} and β_{1j}^{m} are both random, their covariance is needed but cannot be directly estimated; Kenny et al., 2003). Bauer et al. (2006) extended this approach by fitting a multilevel mediation model as a multivariate multilevel model to simultaneously estimate all parameters relevant for estimating and testing indirect effects.

Bauer et al.'s (2006) approach provided more accurate estimates of the indirect effect and its CI. However, it conflated the within-group ("within") component of the indirect effect (the effect involving only individual differences within clusters) and the between-group ("between") component of the indirect effect (the effect involving only cluster means). It is possible, for instance, that the within indirect effect of employee autonomy on productivity through employee motivation is substantial *but* the between indirect effect is negligible. Zhang, Zyphur, and Preacher (2009) addressed this limitation by employing group-mean centering for Level-1 variables (which separates Level-1 variables into level-specific components in a data management step, prior to modeling). Separately estimating within- and between-group coefficients in this way allows for investigation of indirect effects at the group and individual levels for Level-1 variables. However, this method still has a notable limitation in that it can produce biased estimates of the between component of the indirect effect when group means are unreliable (Lüdtke et al., 2008). In addition, a more general limitation of MLM is the inability to model outcomes above the lowest level of the data hierarchy (Krull & MacKinnon, 2001). That is, the effects of individual-level variables on group-level variables, such as in 1-1-2 or 1-2-2 designs, cannot be assessed using MLM. These "bottom-up" effects occur often in group process research and can be of substantive interest (Croon & van Veldhoven, 2007; Kozlowski, Chao, Grand, Braun, & Kuljanin, 2013).

Mediation Within a Multilevel SEM Framework With Fully Nested Data

Limitations of the previous methods can be overcome by extending MLM to include aspects of structural equation modeling (SEM) to yield *multilevel structural equation modeling* (MSEM), of which the aforementioned MLM is a special case. MSEM can be used to produce unbiased estimates of the between indirect effect by treating group means as *latent* variables (Lüdtke et al., 2008; Preacher, Zyphur, & Zhang, 2010). MSEM provides a flexible framework in which many complex relationships among latent and observed variables can be modeled (e.g., Preacher et al., 2010); however, for simplicity we will consider the case where all variables are endogenous.

In MSEM, observed variables can be decomposed into latent within-group and between-group components that may vary within and across groups, respectively:

$$y_{ij} = \tilde{y}_{ij} + \tilde{y}_j \tag{7}$$

$$m_{ij} = \tilde{m}_{ij} + \tilde{m}_j \tag{8}$$

$$x_{ij} = \tilde{x}_{ij} + \tilde{x}_j \tag{9}$$

Here, for individual i in group j, y_{ij} is, as before, the observed dependent variable, m_{ij} is the observed mediator, and x_{ij} is the observed independent variable. $\tilde{y}_{ij}$, $\tilde{m}_{ij}$, and $\tilde{x}_{ij}$ are the latent within-group components of y_{ij}, m_{ij}, and x_{ij}, respectively, and $\tilde{y}_j$, $\tilde{m}_j$, and $\tilde{x}_j$ are the latent between-group components. Coefficients of the within-group model can be treated as random variables that may vary across groups. The within-group model for a 1-1-1 design can be expressed as:

$$\tilde{y}_{ij} = b^y_{1W}\tilde{x}_{ij} + b^y_{2W}\tilde{m}_{ij} + \varepsilon^y_{ij} \tag{10}$$

$$\tilde{m}_{ij} = b^m_{1W}\tilde{x}_{ij} + \varepsilon^m_{ij} \tag{11}$$

$$\tilde{x}_{ij} = \varepsilon^x_{ij} \tag{12}$$

$$\text{where: } \begin{bmatrix} \varepsilon^y_{ij} \\ \varepsilon^m_{ij} \\ \varepsilon^x_{ij} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \theta^y_\varepsilon & & \\ 0 & \theta^m_\varepsilon & \\ 0 & 0 & \theta^x_\varepsilon \end{bmatrix}\right)$$

As before, superscripts for each coefficient and parameter denote the respective outcome variable, and now subscripts B and W denote between- and within-group effects, respectively. b^y_{1W} is the within effect of mediator $\tilde{m}_{ij}$ on $\tilde{y}_{ij}$ controlling for x_{ij} and b^y_{2W} is the within effect of $\tilde{x}_{ij}$ on $\tilde{y}_{ij}$ controlling for m_{ij}. ε^y_{ij} is the within-group residual associated with $\tilde{y}_{ij}$, which is normally distributed with a mean of 0. The estimate b^m_{1W} is the within-group effect of $\tilde{x}_{ij}$ on $\tilde{m}_{ij}$, and ε^m_{ij} and ε^x_{ij} are within-group residuals associated with $\tilde{m}_{ij}$ and $\tilde{x}_{ij}$, respectively.

Similarly, the between-group model for a 1-1-1 design can be expressed as:

$$\tilde{y}_j = b_0^y + b_{1B}^y \tilde{x}_j + b_{2B}^y \tilde{m}_j + \zeta_j^y \tag{13}$$

$$\tilde{m}_j = b_0^m + b_{1B}^m \tilde{x}_j + \zeta_j^m \tag{14}$$

$$\tilde{x}_j = b_0^x + \zeta_j^x \tag{15}$$

where: $$\begin{bmatrix} \zeta_j^y \\ \zeta_j^m \\ \zeta_j^x \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi^y & & \\ 0 & \psi^m & \\ 0 & 0 & \psi^x \end{bmatrix} \right)$$

b_0^y is the intercept for $\tilde{y}_j$, b_{2B}^y is the between-group effect of the mediator $\tilde{m}_j$ on $\tilde{y}_j$ controlling for x_j, b_{1B}^y is the between-group effect of $\tilde{x}_j$ on $\tilde{y}_j$ controlling for m_j, and ζ_j^y is $\tilde{y}_j$'s group-level residual. b_0^m is the intercept for $\tilde{m}_j$, b_{1B}^m is the effect of $\tilde{x}_j$ on $\tilde{m}_j$, and ζ_j^m is $\tilde{m}_j$'s group-level residual. b_0^x is the intercept for $\tilde{x}_j$, and ζ_j^x is its group-level residual.

The MSEM specification in Equations 7–15 for a 1-1-1 design allows testing both within-cluster mediation (by testing $b_{1W}^m b_{2W}^y$) and between-cluster mediation (by testing $b_{1B}^m b_{2B}^y$). A 1-1-1 design is the only three-variable design that permits testing both of these indirect effects. Any other design (e.g., 2-1-1 or 2-1-2) permits testing only between-cluster mediation. MSEM can substantially reduce bias in between-cluster indirect effects compared to MLM (particularly for higher ICC, more groups, and larger group sizes; Preacher, Zhang, & Zyphur, 2011).

Mediation Within a Multilevel SEM Framework With Partially Nested Data

As mentioned earlier, it is also common for study designs in the group processes literature to be partially nested, rather than fully nested, such that clustering is present in one or more study arms, but not other arms (Kirschner et al., 2011; Vadasy & Sanders, 2008). In partially nested designs individuals are typically randomly assigned to condition and then clusters are constructed in one arm. The unclustered arm can be conceptualized as consisting exclusively of clusters of size 1, as we do in what follows. Partially nested designs usually involve different model-implied variances in the clustered versus unclustered arms, and require accounting for both between- and within-cluster variation in the clustered arm.

Models for partially nested designs were first developed in a multilevel modeling framework (MLM-PN; Bauer, Sterba, & Hallfors, 2008; Lee & Thompson, 2005; Moerbeek & Wong, 2008; Roberts & Roberts, 2005). Subsequently, models for partially nested designs were developed in a multivariate SEM and an MSEM framework (Sterba et al., 2014). Here we describe the MSEM specification for partial nesting (MSEM-PN), which has not before been presented focusing on mediation. MSEM-PN uses a *multiple-arm* specification (i.e., a multiple-group specification where *study arm*—perhaps treatment vs. control—is the grouping variable). Below, a *c* superscript denotes parameters and latent components in the clustered arm and a *u* superscript denotes these quantities in the unclustered arm. Two arms are shown here, although there could be more (see Sterba et al., 2014).

The specification of an MSEM-PN for mediation will differ depending on whether *study arm* is a predictor in the mediation pathway or is another (potentially) moderating variable. We focus on the second possibility here and address the first later.

In the clustered arm, between-cluster residual variance is represented by random intercept variances (ψ^{yc}, ψ^{mc}, ψ^{xc}) after accounting for predictors. (Random slopes, while possible to include, are not shown here.) In the clustered arm, within-cluster residual variance is represented by θ_ε^{yc}, θ_ε^{mc}, and θ_ε^{xc}, after accounting for predictors. In the *un*clustered arm, we have the option of estimating residual variance in *either* the between or within model, not both; here we choose the former: ψ^{yu}, ψ^{mu}, ψ^{xu}. Variance components are independent across arms, by design.

Clustered arm:	Unclustered arm:

$$y_{ij} = \tilde{y}_{ij}^{c} + \tilde{y}_{j}^{c} \qquad y_{ij} = \tilde{y}_{ij}^{u} + \tilde{y}_{j}^{u}$$
$$m_{ij} = \tilde{m}_{ij}^{c} + \tilde{m}_{j}^{c} \qquad m_{ij} = \tilde{m}_{ij}^{u} + \tilde{m}_{j}^{u} \tag{16}$$
$$x_{ij} = \tilde{x}_{ij}^{c} + \tilde{x}_{j}^{c} \qquad x_{ij} = \tilde{x}_{ij}^{u} + \tilde{x}_{j}^{u}$$

Between: (17)

$$\tilde{y}_{j}^{c} = b_{0}^{yc} + b_{1B}^{yc}\tilde{x}_{j}^{c} + b_{2B}^{yc}\tilde{m}_{j}^{c} + \zeta_{j}^{yc} \qquad \tilde{y}_{j}^{u} = b_{0}^{yu} + b_{1B}^{yu}\tilde{x}_{j}^{u} + b_{2B}^{yu}\tilde{m}_{j}^{u} + \zeta_{j}^{yu}$$
$$\tilde{m}_{j}^{c} = b_{0}^{mc} + b_{1B}^{mc}\tilde{x}_{j}^{c} + \zeta_{j}^{mc} \qquad \tilde{m}_{j}^{u} = b_{0}^{mu} + b_{1B}^{mu}\tilde{x}_{j}^{u} + \zeta_{j}^{mu}$$
$$\tilde{x}_{j}^{c} = b_{0}^{xc} + \zeta_{j}^{xc} \qquad \tilde{x}_{j}^{u} = b_{0}^{xu} + \zeta_{j}^{xu}$$

Within: (18)

$$\tilde{y}_{ij}^{c} = b_{1W}^{yc}\tilde{x}_{ij}^{c} + b_{2W}^{yc}\tilde{m}_{ij}^{c} + \varepsilon_{ij}^{yc} \qquad \tilde{y}_{ij}^{u} = 0$$
$$\tilde{m}_{ij}^{c} = b_{1W}^{mc}\tilde{x}_{ij}^{c} + \varepsilon_{ij}^{mc} \qquad \tilde{m}_{ij}^{u} = 0$$
$$\tilde{x}_{ij}^{c} = \varepsilon_{ij}^{xc} \qquad \tilde{x}_{ij}^{u} = 0$$

where:

$$\begin{bmatrix} \zeta_{j}^{yc} \\ \zeta_{j}^{mc} \\ \zeta_{j}^{xc} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi^{yc} & & \\ 0 & \psi^{mc} & \\ 0 & 0 & \psi^{xc} \end{bmatrix} \right) \qquad \begin{bmatrix} \zeta_{j}^{yu} \\ \zeta_{j}^{mu} \\ \zeta_{j}^{xu} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi^{yu} & & \\ 0 & \psi^{mu} & \\ 0 & 0 & \psi^{xu} \end{bmatrix} \right)$$

$$\begin{bmatrix} \varepsilon_{ij}^{yc} \\ \varepsilon_{ij}^{mc} \\ \varepsilon_{ij}^{xc} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \theta_{\varepsilon}^{yc} & & \\ 0 & \theta_{\varepsilon}^{mc} & \\ 0 & 0 & \theta_{\varepsilon}^{xc} \end{bmatrix} \right)$$

Partially nested designs may necessitate decomposing effects—including indirect effects—into between and within components in the clustered arm (denoted by the *B* and *W* coefficient subscripts in the clustered arm), but not in the unclustered arm. That is, a 1-1-1 design could be used to test either between-cluster mediation (by testing $b_{1B}^{mc}b_{2B}^{yc}$) and/or within-cluster mediation (by testing $b_{1W}^{mc}b_{2W}^{yc}$) in a *clustered* arm, and could be used to test simple mediation (by testing $b_{1B}^{mu}b_{2B}^{yu}$) in an *unclustered* arm. If researchers want to test the equality of an indirect effect across arms in a partially nested design, they have several options. They could test the equality of a simple indirect effect in the unclustered arm, $b_{1B}^{mu}b_{2B}^{yu}$, to a (a) between indirect effect, $b_{1B}^{mc}b_{2B}^{yc}$, (b) within indirect effect, $b_{1W}^{mc}b_{2W}^{yc}$, or (c) total indirect effect (obtainable by constraining $b_{1B}^{mc} = b_{1W}^{mc}$ and $b_{2B}^{yc} = b_{2W}^{yc}$ and then forming their product) in the clustered arm.

Because the previous MSEM-PN specification is for a 1-1-1 design, it does not contain Level-2 predictors or outcomes. More generally, however, in the clustered arm we can distinguish between Level-1 and Level-2 predictors and/or outcomes. But in the unclustered arm, variables do not have an inherent level. Variables measured at Level-1 in the clustered arm are also usually measured in the unclustered arm (as in Equation 16). However, variables measured at Level-2 in the clustered arm may be either missing-by-design[1] or measured in the unclustered arm. For instance, a Level-2 variable measured for teams in the clustered arm but missing-by-design for individuals in the unclustered arm could be *team closeness*. But a variable measured for teams in the

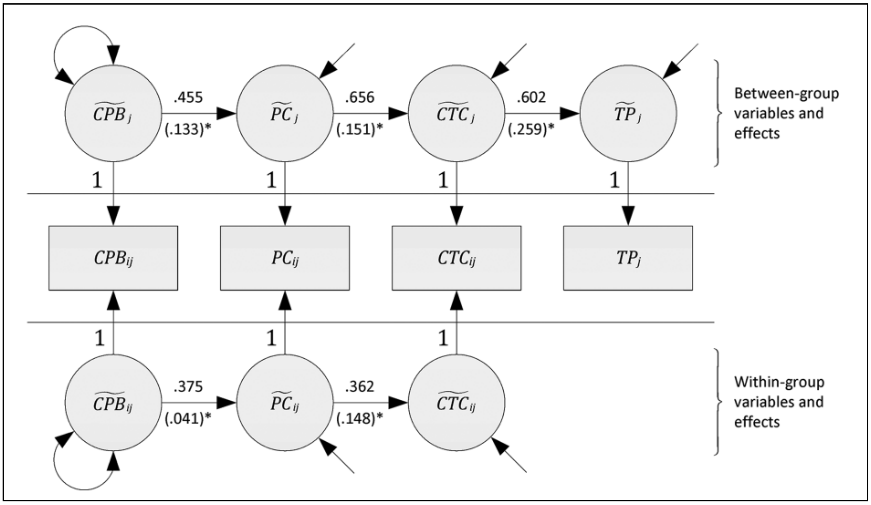

Figure 1. Illustration of a MSEM model for a fully nested 1-1-1-2 design (Example 1).
Note. This MSEM is patterned after the model of Nohe et al. (2013). *CPB* = team leader-rated change-promoting behavior; *PC* = team member-rated perceived charisma; *CTC* = team member-rated commitment to change; *TP* = team leader-rated team performance. Circles are latent within or between components of measured variables (squares). Straight arrows that are labeled with estimates are regression paths. Curved arrows are residual variances. Path coefficients are unstandardized. Numbers in parentheses are standard errors.
$^*p < .05$.

clustered arm and for individuals in the unclustered arm could be *problem-solving speed.* If all Level-2 variables in the clustered arm are *also* measured in the unclustered arm, a researcher with a 2-2-2, 2-1-1, 2-1-2, 1-2-1, 2-2-1, 1-1-2, or 1-2-2 design could test between-cluster mediation in the clustered arm *and* simple mediation in the unclustered arm. If, on the other hand, Level-2 variables in the clustered arm are missing-by-design in the unclustered arm, then for these seven designs *no* mediation could be tested in the unclustered arm.

Note that mediation analyses using MSEM-PN do not require that *study arm* be a potential moderator of the mediation pathway, as in Equations 16–18. Rather, *study arm* could replace x_{ij} as a (now, Level-2) predictor variable in the mediation pathway, while still keeping within the multiple-arm framework. To achieve this, equations with $\tilde{x}^c_j$, $\tilde{x}^c_{ij}$, or $\tilde{x}^u_j$ as outcomes would drop out of Equations 16–18, as would slopes of $\tilde{x}^c_j$, $\tilde{x}^c_{ij}$, or $\tilde{x}^u_j$, and we would constrain $b^{yc}_{2B} = b^{yu}_{2B} = b^{y}_{2B}$. In this model, the between-cluster indirect effect of $\tilde{x}_j$ on $\tilde{y}_j$ through $\tilde{m}_j$, for instance, could still be obtained by multiplying $(b^{mc}_0 - b^{mu}_0) \times (b^{y}_{2B})$ (whereas the between direct effect could be obtained as $[b^{yc}_0 - b^{yu}_0]$).

It should be noted that the partial nesting design has a unique limitation for internal validity and unique strength for external validity, as described by Bauer et al. (2008). Specifically, internal validity is limited by the fact that treatment and grouping effects are conflated; we do not know the result of merely grouping participants without administering treatment. However, external validity is strengthened by the fact that the unclustered control arm may more accurately reflect real-world conditions occurring in the absence of treatment.

Examples

To illustrate how MSEM is used to assess mediation in clustered data, we simulated a fully nested dataset and a partially nested dataset with data-generating parameters based primarily on a fully

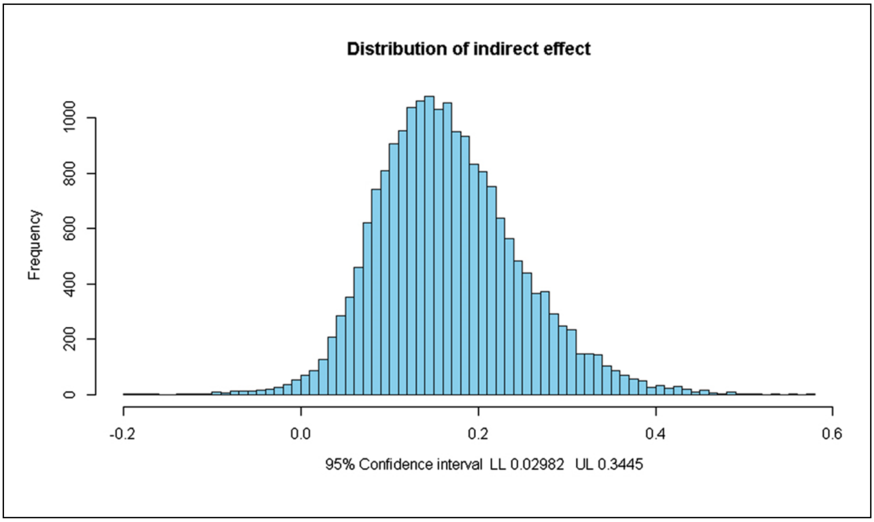

Figure 2. Sampling distribution of the between-cluster indirect effect of $\widetilde{CPB}_j$ on $\widetilde{TP}_j$ through $\widetilde{PC}_j$ and $\widetilde{CTC}_j$ from the fully nested MSEM (Example 1).
Note. Distribution created in R (Version 3.0.3).

nested design from a study by Nohe et al. (2013)[2] where the Level-1 unit is the individual worker and Level-2 unit is the team. This simulated example involves a 1-1-1-2 design, which involves two new complexities not discussed earlier: two mediators (not one) and a combination of Level-2 and Level-1 outcomes. Similarly to Nohe et al.'s original analysis, we are interested in determining whether a leader's perceived change-promoting behaviors (CPB_{ij}) affected team performance (TP_j) through perceived leader charisma (PC_{ij}) and commitment to change (CTC_{ij}; see Figure 1 and online Appendix for the full set of corresponding equations). Team performance is a Level-2 variable, and perceived change-promoting behavior, perceived charisma, and commitment to change are Level-1 variables.

Example 1: Fully Nested 1-1-1-2 Design

More specifically, we are interested in estimating the *between-cluster indirect effect* of $\widetilde{CPB}_j$ on $\widetilde{TP}_j$ through (first) $\widetilde{PC}_j$ and (second) $\widetilde{CTC}_j$. We are also interested in decomposing the indirect effect of CPB_{ij} on CTC_{ij} through PC_{ij} into *between-cluster* and *within-cluster* indirect effects, and testing their equality. Note that in Figure 1 not all direct effects are estimated, following Nohe et al. (2013). MSEM is beneficial to use here because (a) if mediators PC_{ij} and CTC_{ij} were not split into *latent* between and within components, any between indirect effects involving them would be biased, and (b) MSEM allows estimation of the path linking $\widetilde{CTC}_j$ to $\widetilde{TP}_j$.

In the original study, participants were 33 teams ranging in size from two to 10 members, yielding a total of 142 team members from a large German company. CPB_{ij} refers to how members rate their leader's engagement with change-promoting activities. In Nohe et al. (2013), PC_{ij} was assessed using three items that asked followers to rate leader charisma (e.g., "My leader acts in ways that build my respect"). CTC_{ij} was assessed using four items that asked followers to rate their commitment to change (e.g., "This change serves an important purpose"). TP_j was assessed using four items that asked leaders to rate team performance (e.g., "Accomplishes most of their tasks quickly and efficiently"). Using M*plus* (v. 7.11; Muthén & Muthén, 1998–2014) we simulated a data set using the authors' reported parameter estimates as population parameters,[3] and we assumed all residuals and random effects were normally distributed. We generated data to

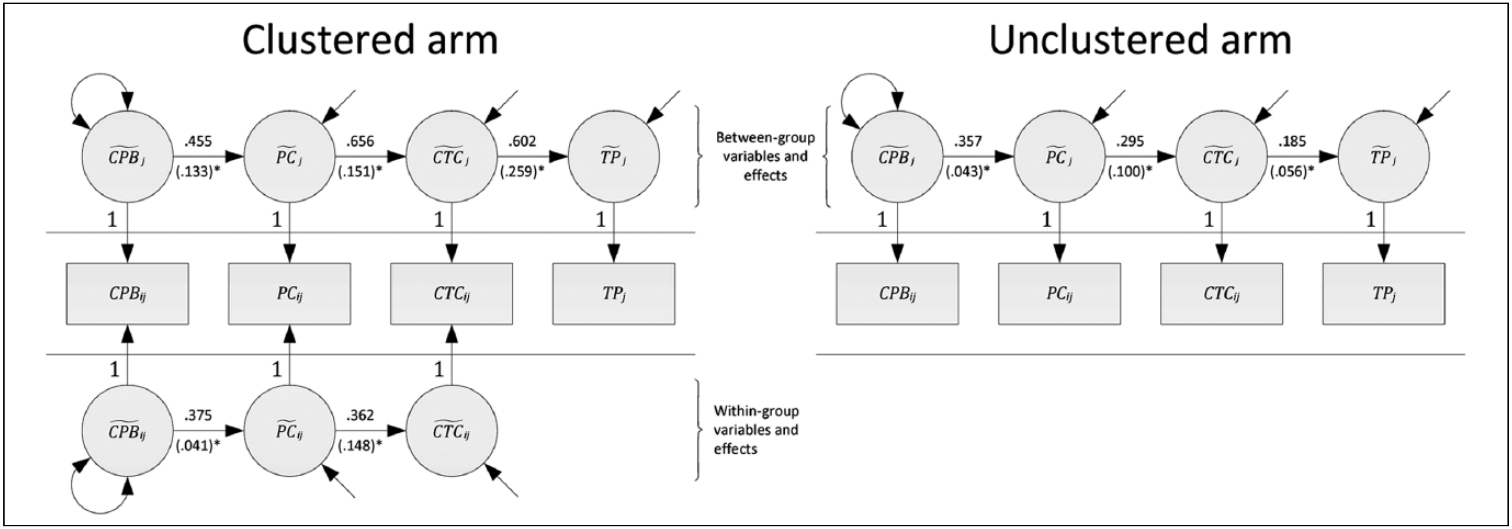

Figure 3. Illustration of a MSEM-PN model for a partially nested 1-1-1-2 design (Example 2).
Note. In the unclustered arm, each person is considered to be their own cluster. *CPB* = team leader-rated change-promoting behavior; *PC* = team member-rated perceived charisma; *CTC* = team member-rated commitment to change; *TP* = team leader-rated team performance. Path coefficients are unstandardized. Circles are latent within or between components of measured variables (squares). Straight arrows that are labeled with estimates and standard errors in parentheses are regression paths. Curved arrows are residual variances.
$^*p < .05$.

consist of 142 members nested within 33 teams, as in the original study; see the online Appendix for the data.

This data set was analyzed using MSEM; see the Appendix for annotated M*plus* analysis syntax and the online Appendix for annotated output. Specifying the analysis type as "TWOLEVEL" allows for estimation of within- and between-team components. The "%WITHIN%" section contains a model for the within components of Level-1 variables (CPB_{ij}, PC_{ij}, and CTC_{ij}), including within-team residual variances and path coefficients; the "%BETWEEN%" section contains a model for the Level-2 variables and the between components of Level-1 variables. The between indirect effects and within indirect effect of interest were computed in the "MODEL CONSTRAINT" section.

Results. Results are reported in Figure 1. ICCs for CPB_{ij}, PC_{ij}, and CTC_{ij} were .510, .587, and .196, respectively. A team leader's $\widetilde{CPB}_j$ increased average individual ratings of $\widetilde{PC}_j$ at the between-team level (.455, p = .001). The change in $\widetilde{PC}_j$ was then associated with increased average individual ratings of $\widetilde{CTC}_j$ at the between-team level (.656, p < .001), which, in turn, resulted in improved team $\widetilde{TP}_j$ (.602, p = .020). In addition, a team leader's $\widetilde{CPB}_{ij}$ increased ratings of $\widetilde{PC}_{ij}$ at the individual level (.375, p < .001). The change in $\widetilde{PC}_{ij}$ was then associated with increased ratings of $\widetilde{CTC}_{ij}$ at the individual level (.362, p = .015).

The indirect effect of CPB_{ij} on CTC_{ij} through PC_{ij} was split into a between-cluster indirect effect (.299, CI = {.097, .574}) and a within-cluster indirect effect (.136, CI = {.027, .252}) which were not significantly different (difference = .163, CI = {−.084, .464}). Because the indirect effect of $\widetilde{CPB}_j$ on $\widetilde{TP}_j$ through $\widetilde{PC}_j$ and $\widetilde{CTC}_j$ terminates with a Level-2 variable, this indirect effect can exist only at the between-team level (.18, CI = {.030, .345}; Zhang et al., 2009). Because these three indirect effects are each a product of normally distributed regression coefficients, their distributions are nonnormal (see Figure 2) and require asymmetric CIs. The CIs reported here are Monte Carlo 95% confidence intervals (Preacher & Selig, 2012); computational details are provided in the online Appendix. Because the 95% CIs for the between- and within-cluster indirect effect of CPB_{ij} on CTC_{ij} through PC_{ij} and the between-cluster indirect effect of $\widetilde{CPB}_j$ on $\widetilde{TP}_j$ through $\widetilde{PC}_j$ and $\widetilde{CTC}_j$ did not contain zero, these indirect effects were significant at α = .05. However, the difference between the within- and between-cluster indirect effects of CPB_{ij} on CTC_{ij} through PC_{ij} was nonsignificant. These results, using simulated data, indicate that the relationship between team members' average perception of change-promoting behavior and team performance was mediated by team members' average perceived charisma of the leader and the team members' average commitment to change.

Example 2: Partially Nested 1-1-1-2 Design

In this example, we considered the generated data from Example 1 to constitute the clustered arm of a two-arm study. To form a partially nested data set, we generated data from a second (unclustered) arm of 142 individuals; these individuals work independently rather than in teams. In this unclustered arm of the generating MSEM-PN for this 1-1-1-2 design, CPB_{ij} again affected TP_j through PC_{ij} and CTC_{ij}. Even though TP_j was measured at Level-2 in the clustered arm, it was also measured in the unclustered arm, where each individual constitutes his or her own team.[4] Across arms in the population, all intercepts were equal, and the residual variances for $\widetilde{CPB}_j$, $\widetilde{PC}_j$, and $\widetilde{CTC}_j$ in the unclustered arm were equal to their total residual variances in the clustered arm. All slopes in the unclustered arm were equal in the population to the between effects from the clustered arm, except the effect of CTC_{ij} on TP_j (now .25). The residual variance of TP_j is now .377.

We fit the generating MSEM-PN model to this 1-1-1-2 data set; see Appendix A for M*plus* syntax for model fitting and see the online Appendix for model equations and output. Research questions of interest might involve testing the three indirect effects assessed in Example

1 in the clustered arm, as well as testing two indirect effects in the unclustered arm: the four-variable simple indirect effect of $\widetilde{CPB}_j$ on $\widetilde{TP}_j$ through $\widetilde{PC}_j$ and $\widetilde{CTC}_j$, and the three-variable simple indirect effect of $\widetilde{CPB}_j$ on $\widetilde{CTC}_j$ through $\widetilde{PC}_j$. Also, of key interest here is testing the equality of simple indirect effects in the unclustered arm with particular indirect effects in the clustered arm.

Results. See Figure 3 for full results; selected results are described here. Similarly to the clustered arm, in the unclustered arm both the three-variable mediation relationship (simple indirect effect = .105, CI = {.034, .183}) and the four-variable mediation relationship (simple indirect effect = .019, CI = {.005, .040}) differed significantly from 0. Furthermore, the four-variable simple indirect effect of $\widetilde{CPB}_j$ on $\widetilde{TP}_j$ through $\widetilde{PC}_j$ and $\widetilde{CTC}_j$ for individual workers in the unclustered arm was found to differ significantly *across arm* from its counterpart between-team indirect effect in the clustered arm. The latter was larger by .16, CI = {.009, .327}. This implies that at least some part of the mechanism whereby perceived leader characteristics affect motivational outcomes of workers (i.e., average commitment) and behavioral outcomes of workers (i.e., average task performance) operates differently for multiperson teams than for individual workers (i.e., singleton teams). In practice, substantive context would determine whether to compare a simple indirect effect in the unclustered arm to a between indirect effect in the clustered arm (as done here) or to a simple or total indirect effect in the clustered arm.[5]

Discussion

The purposes of this article were threefold. The first was to highlight the problems that arise if clustering is ignored when estimating indirect effects in fully nested or partially nested data. We also briefly reviewed strengths and weaknesses of MLM methods that have been employed to address clustering when assessing mediation.

The second purpose of this article was to highlight the advantages of MSEM for assessing mediation in clustered data. MSEM can more accurately partition the variance of between- and within-group components of multilevel data, especially when mediation occurs at the between-group level. In addition, MSEM allows estimation of effects of individual-level predictors on group-level outcomes (i.e., bottom-up effects). Finally, it should also be noted that MLM models can be considered special cases of MSEM, making MSEM a more general and flexible framework in which to consider multilevel mediation (Preacher et al., 2011; Preacher et al., 2010). MSEM-PN also provides a new flexible approach for testing mediation in partially nested designs where it is possible to assess between- and/or within-cluster mediation in a clustered arm, and simple mediation in an unclustered arm. Options for testing the equality of certain indirect effects across clustered and unclustered arms were introduced here.

The third purpose of this article was to demonstrate how to implement existing MSEM analyses and new MSEM-PN specifications for nested and partially nested data, respectively, using illustrative examples. We provided instructions and example M*plus* syntax for fitting MSEM and MSEM-PN to a 1-1-1-2 design and for calculating between- and/or within-cluster indirect effects. We tested the significance of indirect effects of interest using a Monte Carlo CI procedure (Preacher & Selig, 2012).

Extensions

The data in the previous examples included continuous mediators and outcomes from (up to) two levels. However, it is possible to model clustering at higher and/or lower levels of the data hierarchy. For example, workers could be (fully or partially) nested in teams, which are fully nested within corporations, or repeated measures could be fully nested within workers which are (fully or partially) nested in teams. Additionally, there could be categorical mediators, and/or outcome variables; a categorical outcome could include, at

Level 1, whether an individual leaves a company or, at Level 2, successful completion of a team task.

Recommendations for Implementation

We conclude with recommendations for researchers who plan to implement an MSEM or MSEM-PN analysis. Before collecting data, researchers should identify a theoretically justified mediation model and posit at which levels of the data hierarchy mediation is expected to occur. Between-group variability and sample size should also be considered when planning to employ MSEM. A low proportion of between-group variability (i.e., ICC < .05) can lead to unstable parameter estimates or lack of model convergence. Based on simulation study results, Preacher et al. (2011) recommended that ideally group sizes be at least 20 when ICCs were small (i.e., ICC < .05), and, in general, demonstrated that increasing group sizes, number of groups, and ICCs improved the stability and accuracy of parameter estimates. Additional guidance on choosing Level-1 and Level-2 sample sizes when using MSEM can be found in Li and Beretvas (2013).

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. Sterba et al. (2014) provide a data management procedure that can avoid listwise deletion of exogenous predictors that are missing-by-design in the unclustered arm when a conditional likelihood is used for model fitting. Our use of endogenous predictors here avoids listwise deletion without employing this procedure.
2. Nohe et al. (2013) used a 2-1-1-2 mediation design but we modified it to a 1-1-1-2 design here for pedagogical purposes.
3. The effect of $\widetilde{CPB}_{ij}$ on $\widetilde{PC}_{ij}$ (not in Nohe et al.'s [2013] original analysis) was generated to be the same as the existing effect of $\widetilde{CPB}_j$ on $\widetilde{PC}_j$ in the population. The within variance of CPB_{ij} was .178.
4. An example in which a Level-2 outcome is missing-by-design in an unclustered arm of a partially nested design is given in Sterba et al. (2014).
5. For instance, if a researcher were instead interested in testing the equality of the three-variable indirect effect of $\widetilde{CPB}_j$ on $\widetilde{CTC}_j$ through $\widetilde{PC}_j$ in the unclustered arm to its counterpart within-team indirect effect in the clustered arm ($\widetilde{CPB}_{ij}$ on $\widetilde{CTC}_{ij}$ through $\widetilde{PC}_{ij}$), this difference would be nonsignificant here (difference = .03, CI = {–.103, .167}).

References

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*, 142–163. doi:10.1037/1082–989X.11.2.142

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research*, *43*, 210–236. doi:10.1080/00273170802034810

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, *20*, 115–140. doi:10.2307/271084

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

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

Kirschner, F., Paas, F., Kirschner, P. A., & Janssen, J. (2011). Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learning and Instruction*, *21*, 587–599. doi:10.1016/j.learninstruc.2011.01.001

Kozlowski, S. J., Chao, G. T., Grand, J. A., Braun, M. T., & Kuljanin, G. (2013). Advancing multilevel research design: Capturing the dynamics of emergence. *Organizational Research Methods*, *16*, 581–615. doi:10.1177/1094428113493119

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

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

Lee, K. J., & Thompson, S. G. (2005). The use of random effects models to allow for clustering in individually randomized trials. *Clinical Trials*, *2*, 163–173. doi:10.1191/1740774505cn082oa

Li, X., & Beretvas, S. N. (2013). Sample size limits for estimating upper level mediation models using multilevel SEM. *Structural Equation Modeling*, *20*, 241–264. doi:10.1080/10705511.2013.769391

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

MacKinnon, D. P., Fritz, M. S., Williams, J. C., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, *39*, 384–389. doi:10.3758/BF03193007

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

Moerbeek, M., & Wong, W. K. (2008). Sample size formulae for trials comparing group and individual treatments in a multilevel model. *Statistics in Medicine*, *27*, 2850–2864. doi:10.1002/sim.3115

Muthén, L. K., & Muthén, B. (1998–2014). *Mplus user's guide*. (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nohe, C., Michaelis, B., Menges, J. I., Zhang, Z., & Sonntag, K. (2013). Charisma and organizational change: A multilevel study of perceived charisma, commitment to change, and team performance. *The Leadership Quarterly*, *24*, 378–389. doi:10.1016/j.leaqua.2013.02.001

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. doi:10.1080/00273170802034844

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

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

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*, 77–98. doi:10.1080/15427600902911247

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

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Roberts, C., & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials*, *2*, 152–162. doi:10.1191/1740774505cn076oa

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422–445. doi:10.1037//1082-989x.7.4.422

Sterba, S. K., Preacher, K. J., Hardcastle, E. J., Forehand, R., Cole, D. A., & Compas, B. E. (2014). Structural equation modeling approaches for analyzing partially nested data. *Multivariate Behavioral Research*, *49*, 93–118. doi:10.1080/00273171.2014.882253

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

Vadasy, P. F., & Sanders, E. A. (2008). Benefits of repeated reading intervention for low-achieving fourth- and fifth-grade students. *Remedial and Special Education*, *29*, 235–249. doi:10.1177/0741932507312013

Van Mierlo, H., Rutte, C. G., Vermunt, J. K., Kompier, M. A. J., & Doorewaard, J. A. C. M. (2007). A multi-level mediation model of the relationships between team autonomy, individual task design and psychological well-being. *Journal of Occupational* *and Organizational Psychology*, *80*, 647–664. doi:10.1348/096317907X196886

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*, 301–322. doi:10.1037/a0016972

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12*, 695–719. doi:10.1177/1094428108327450

Zhou, L., Wang, M., Chen, G., & Shi, J. (2012). Supervisors' upward exchange relationships and subordinate outcomes: Testing the multilevel mediation role of empowerment. *Journal of Applied Psychology*, *97*, 668–680. doi:10.1037/a0026305

Appendix A

MSEM syntax (fully nested 1-1-1-2 example 1 mediation model).

```
DATA: FILE IS fullynest_seed100.dat; !call dataset
VARIABLE:
NAMES ARE TP PC CTC CPB CLUSTER; !name variables
USEVARIABLES ARE TP PC CTC CPB; !variables to be used in analysis
CLUSTER=CLUSTER; !identify clustering variable
BETWEEN ARE TP; !identify between-cluster (level-2) variables
ANALYSIS: TYPE IS TWOLEVEL;
MODEL:
%WITHIN% !within-cluster (level-1) model
PC CTC CPB; !estimate within-cluster variances
CTC ON PC (bw); !regress CTC on PC, call the slope "bw"
PC ON CPB (aw); !regress PC on CPB, call the slope "aw"
%BETWEEN% !between-group (level-2) model
CPB PC CTC TP; !estimate between-cluster variances
PC ON CPB (ab); !regress PC on CPB, call the slope "ab"
CTC ON PC (bb); !regress CTC on PC, call the slope "bb"
TP ON CTC (cb); !regress CTC on TP, call the slope "cb"
[CPB TP PC CTC]; !estimate variable means

MODEL CONSTRAINT:
NEW(indb_3 indb_2 indw_2 diff); !create indirect effect parameters
indb_3=ab*bb*cb; !between-cluster indirect effect 1
indb_2=ab*bb; !between-cluster indirect effect 2
indw_2=aw*bw; !within-cluster indirect effect
diff=indb_2-indw_2; !diff. of within- and between-cluster indirect
effects

OUTPUT: TECH1 TECH3 SVALUES;
```

MSEM -PN syntax (partially nested 1-1-1-2 example 2 mediation model)

```
DATA: FILE IS bothnest_seed100.dat;
VARIANCES=NOCHECK;

VARIABLE: NAMES ARE TP PC CTC CPB CLUSTER treat; !name variables
USEVARIABLES ARE TP PC CTC CPB; !identify variables for analysis
CLUSTER IS cluster; !identify clustering variable
GROUPING IS treat (0=cont 1=txt); !identify grouping (study arm) variable
BETWEEN IS tp; !identify between-cluster (level-2) variables
ANALYSIS: TYPE IS TWOLEVEL;

MODEL: !model for tx (nested) group
%WITHIN% !within-cluster (level-1) model for tx group
PC CTC CPB; !estimate within-cluster variances
CTC ON PC (bw); !regress CTC on PC, call the slope "bw"
PC ON CPB (aw); !regress PC on CPB, call the slope "aw"

%BETWEEN% !between-cluster (level-2) model for tx group
CPB PC CTC TP; !estimate between-cluster variances
PC ON CPB (ab); !regress PC on CPB, call the slope "ab"
CTC ON PC (bb); !regress CTC on PC, call the slope "bb"
TP ON CTC (cb); !regress TP on CTC, call the slope "cb"
[CPB TP PC CTC]; !estimate means

MODEL cont: !model for control (non-nested) group
%WITHIN% !within-cluster (level-1) model for control group
PC@0; CTC@0; CPB@0; !all variances set to 0
CTC ON PC@0; !regression of CTC on PC set to 0
PC ON CPB@0; !regression of PC on CPB set to 0

%BETWEEN% !between-cluster (level-2) model for control group
CPB PC CTC TP; !estimate between-cluster variances
PC ON CPB (a); !regress PC on CPB, call the slope "a"
CTC ON PC (b); !regress CTC on PC, call the slope "b"
TP ON CTC (c); !regress TP on CTC, call the slope "c"
[CPB TP PC CTC]; !estimate means

MODEL CONSTRAINT:
!create indirect effect variables
NEW(indb_3 indb_2 indw_2 diff inds_3 inds_2 diff3_arm diff2_arm);
indb_3=ab*bb*cb; !first between-cluster indirect effect for tx group
indb_2=ab*bb; !second between-cluster indirect effect for tx group
indw_2=aw*bw; !within-cluster indirect effect for tx group
diff=indb_2-indw_2; !diff. in within- and between-cluster
          !indirect effects for tx group
```

```
inds_3=a*b*c; !first indirect effect for control group
inds_2=a*b; !second indirect effect for control group
diff3_arm=indb_3-inds_3; !diff. in first indirect effect
            !across tx and control groups
diff2_arm=indw_2-inds_2; !diff. in second indirect effect
            !across tx and control groups

OUTPUT: NOCHISQUARE TECH1 TECH3 SVALUES;
```