A General Multilevel SEM Framework for Assessing Multilevel Mediation

Kristopher J. Preacher
University of Kansas

Michael J. Zyphur
University of Melbourne

Zhen Zhang
Arizona State University

Several methods for testing mediation hypotheses with 2-level nested data have been proposed by researchers using a multilevel modeling (MLM) paradigm. However, these MLM approaches do not accommodate mediation pathways with Level-2 outcomes and may produce conflated estimates of between- and within-level components of indirect effects. Moreover, these methods have each appeared in isolation, so a unified framework that integrates the existing methods, as well as new multilevel mediation models, is lacking. Here we show that a multilevel structural equation modeling (MSEM) paradigm can overcome these 2 limitations of mediation analysis with MLM. We present an integrative 2-level MSEM mathematical framework that subsumes new and existing multilevel mediation approaches as special cases. We use several applied examples and accompanying software code to illustrate the flexibility of this framework and to show that different substantive conclusions can be drawn using MSEM versus MLM.

Keywords: multilevel modeling, structural equation modeling, multilevel SEM, mediation

Supplemental materials: http://dx.doi.org/10.1037/a0020141.supp

Researchers in behavioral, educational, and organizational research settings often are interested in testing mediation hypotheses with hierarchically clustered data. For example, Bacharach, Bamberger, and Doveh (2008) investigated the mediating role of distress in the relationship between the intensity of involvement in work-related incidents and problematic drinking behavior among firefighting personnel. They used data in which firefighters were nested within ladder companies and all three variables were assessed at the subject level. Using data from customer service engineers working in teams, Maynard, Mathieu, Marsh, and Ruddy (2007) found that team-level interpersonal processes mediate the relationship between team-level resistance to empowerment and individual job satisfaction. Both of these examples—and many others—involve data that vary both within and between higher level units. Traditional methods for assessing mediation (e.g., Baron & Kenny, 1986; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Warsi, & Dwyer, 1995) are inappropriate in these multilevel settings, primarily because the assumption of independence of observations is violated when clustered data are used, leading to downwardly biased standard errors if ordinary regression is used. For this reason, several methods have been proposed for addressing mediation hypotheses when the data are hierarchically organized.

These recommended procedures for testing multilevel mediation have been developed and framed almost exclusively within the standard multilevel modeling (MLM) paradigm (for thorough treatments of MLM, see Raudenbush & Bryk, 2002, and Snijders & Bosker, 1999) and implemented with commercially available MLM software, such as SAS PROC MIXED, HLM, or MLwiN. For example, some authors have discussed models in which the independent variable X, mediator M, and dependent variable Y all are measured at Level 1 of a two-level hierarchy (a 1-1-1 design, adopting notation proposed by Krull & MacKinnon, 2001),[1] and slopes either are fixed (Pituch, Whittaker, & Stapleton, 2005) or are permitted to vary across Level-2 units (Bauer, Preacher, & Gil, 2006; Kenny, Korchmaros, & Bolger, 2003). Other mediation models have also been examined, including mediation in 2-2-1 designs, in which both X and M are assessed at the group level (Krull & MacKinnon, 2001; Pituch, Stapleton, & Kang, 2006), and mediation in 2-1-1 designs, in which only X is assessed at the group level (Krull & MacKinnon, 1999, 2001; MacKinnon, 2008; Pituch & Stapleton, 2008; Raudenbush & Sampson, 1999).

Each of these approaches from the MLM literature was suggested in response to the need to estimate a particular model to test a mediation hypothesis in a specific design. However, the MLM paradigm is unable to accommodate simultaneous estimation of a

---

[1] We slightly alter the purpose of Krull and MacKinnon's (2001) notation to refer to data collection designs rather than to models. The models we propose for various multilevel data designs are not consistent with models that have appeared in prior research.

number of other three-variable multilevel mediation relationships that can occur in practice (see Table 1; models for some of these designs were also treated by Mathieu & Taylor, 2007). Designs 1 and 2 in Table 1 cannot be accommodated easily or completely in the MLM framework, because multivariate models require significant additional data management (Bauer et al., 2006), and MLM does not fully separate between-group and within-group effects without introducing bias. Designs 3 through 7 in Table 1 cannot be accommodated in the MLM framework at all because of MLM's inability to model effects involving upper level dependent variables.

In this article, we first explain these two reasons in detail and show how they are circumvented by adopting a multilevel structural equation modeling (MSEM) framework for testing multilevel mediation. Because a general framework for multilevel mediation in structural equation modeling (SEM) has yet to be presented, we then introduce MSEM and show how Muthén and Asparouhov's (2008) general MSEM mathematical framework can be applied in investigating multilevel mediation. We show how all previously proposed multilevel mediation models—as well as newly proposed multilevel mediation models—can be subsumed within the MSEM framework. Third, we present several empirical examples (with thoroughly annotated code) in order to introduce and aid in the interpretation of new and former multilevel mediation models that the MSEM framework, but not the MLM framework, can accommodate. The primary contributions of this article are to show (a) how MSEM can separate some variables and effects into within- and between-group components to yield a more thorough and less misleading understanding of indirect effects in hierarchical data and (b) how MSEM can be used to accommodate mediation models containing mediators and outcome variables assessed at Level 2 (e.g., bottom-up effects). We advocate the use of MSEM as a comprehensive system for examining mediation effects in multilevel data.

First Limitation of MLM Framework for Multilevel Mediation: Conflation of Between and Within Effects (Designs 1 to 3 in Table 1)

In two-level designs, it is possible to partition the variance of a variable in clustered data into two orthogonal latent components, the *Between (cluster) component* and the *Within (cluster) component* (Asparouhov & Muthén, 2006). Variables assessed at Level 2 have only Between components of variance. Variables assessed at Level 1 typically have both Between and Within components, although in some cases a Level-1 variable may have only a Within component if it has no between-group variation.[2,3] If a variable has both Between and Within variance components, the Between component is necessarily uncorrelated with the Within component of that variable and the Within components of all other variables in the model. Similarly, the Within component of a variable is necessarily uncorrelated with the Between component of that variable and the Between components of all other variables in the model. We refer to effects of Between components (or variables) on other Between components (or variables) as *Between effects* and to effects of Within components (or variables) on other Within components (or variables) as *Within effects*. Because Between and Within components are uncorrelated, it is not possible for a Between component to affect a Within component or vice versa.

Ordinary applications of multilevel modeling do not distinguish Between effects from Within effects and instead report a single mean slope estimate that combines the two (see, relatedly, Chan, 1998; Cheung & Au, 2005; Klein, Conn, Smith, & Sorra, 2001). This is an important issue for mediation researchers to consider; the use of slopes that combine Between and Within effects can easily lead to indirect effects that are biased relative to their true values, because the component paths may conflate effects that are relevant to mediation with effects that are not.

If the Within effect of a Level-1 predictor is different from the Between effect (i.e., a compositional or contextual effect exists; Raudenbush & Bryk, 2002), the estimation of a single conflated slope ensures that—regardless of the level for which the researcher wishes to make inferences—the resulting slope will mischaracterize the data. The problem of conflated Within and Between effects in MLM is already well recognized for ordinary slopes in multilevel models (e.g., Asparouhov & Muthén, 2006; Hedeker & Gibbons, 2006; Kreft, de Leeuw, & Aiken, 1995; Lüdtke et al., 2008; Mancl, Leroux, & DeRouen, 2000; Neuhaus & Kalbfleisch, 1998; Neuhaus & McCulloch, 2006), but has been underemphasized in literature on multilevel mediation.[4]

To illustrate why this issue is problematic for mediation research, we will consider the case in which X is assessed at Level 2 and M and Y are assessed at Level 1 (i.e., 2-1-1 design). The effect of X on Y must be a strictly Between effect; because X is constant within a given group, variation in X cannot influence individual differences within a group (Hofmann, 2002).[5] Therefore, any mediation of the effect of a Level-2 X must also occur at a between-group level, regardless of the level at which M and Y are assessed, because the only kind of effect that X can exert (whether direct or indirect) must be at the between-group level. In fact, any mediation effect in a model in which at least one of X, M, or Y is assessed at Level 2 must occur strictly at the between-group level.

Now consider the 2-1 portion of the 2-1-1 design. The first component of the indirect effect (a, the effect of X on M) is estimated properly in standard MLM. That is, the effect of X on M and the effect of X on Y are unbiased in the MLM framework. (However, it is not generally recognized that—at an interpreta-

---

[2] The Between/Within variance components distinction pertains to latent sources of variance. This concept is related to, but distinct from, the group mean and grand mean centering choices commonly discussed for observed variables in the traditional MLM literature. In the MSEM approach, observed Level-1 variables are typically either grand mean centered or not centered prior to analysis. In either case, in MSEM all Level-1 variables are subjected to implicit, model-based group mean centering by default unless constraints are applied to the model.

[3] In rare cases it is possible for a Level-1 variable to have a negative estimated Level-2 variance (Kenny, Mannetti, Pierro, Livi, & Kashy, 2002), but we do not discuss such cases here.

[4] Exceptions are Krull and MacKinnon (2001), MacKinnon (2008), and Zhang, Zyphur, and Preacher (2009), who explicitly separated the Within and Between effects in the 1-1 portion of a mediation model for 2-1-1 data.

[5] It is assumed here that, if X is a treatment variable, individuals within a cluster receive the same "dosage" or amount of treatment. If information on dosage or compliance is available, it should be included in the model.

Table 1
Possible Two-Level Multilevel Mediation Designs

No.	Design	Methods proposed	Investigated by	Limitations of MLM
1	2-1-1	MLM	Kenny et al. (1998, 2003); Krull & MacKinnon (1999, 2001); MacKinnon (2008); Pituch & Stapleton (2008); Pituch et al. (2006); Raudenbush & Sampson (1999); Zhang et al. (2009)	Conflation or bias of the indirect effect
2	1-1-1	MLM	Bauer et al. (2006); Kenny et al. (2003); Krull & MacKinnon (2001); MacKinnon (2008); Pituch et al. (2005)	Conflation or bias of the indirect effect
		MSEM	Raykov & Mels (2007)	
3	1-1-2	None	None	MLM cannot be used; Level-2 dependent variables not permitted
4	2-2-1	OLS + MLM	Krull & MacKinnon (2001); Pituch et al. (2006)	Two-step, nonsimultaneous estimation; Level-2 dependent variables not permitted
		SEM	Bauer (2003)	
5	1-2-1	None	None	MLM cannot be used; Level-2 dependent variables not permitted
6	1-2-2	None	None	MLM cannot be used; Level-2 dependent variables not permitted
7	2-1-2	None	None	MLM cannot be used; Level-2 dependent variables not permitted

Note. MLM = traditional multilevel modeling; MSEM = multilevel structural equation modeling; OLS = ordinary least squares; SEM = structural equation modeling.

tional level—these are actually Between effects because *X* may affect only the Between components of *M* and *Y*.)

Now consider the 1-1 component of the 2-1-1 design. The second portion of the indirect effect (*b*, the effect of *M* on *Y*), arguably, is not estimated properly in standard MLM. This effect is separable into two parts, one occurring strictly at the between-group level and the other strictly at the within-group level. The use of slopes that combine between- and within-group effects is known to create substantial bias in multilevel models (Lüdtke et al., 2008). When the true Within effect is greater than the true Between effect, there is an upward bias in the conflated effect. When the true Within effect is less than the true Between effect, downward bias is present. Because the *b* component is part of the Between indirect effect that is of primary interest in 2-1-1 designs, the indirect effect is similarly biased. This is true of any design involving effects linking Level-1 variables (e.g., 1-1-1, 2-1-1, and longer causal chains containing at least a single 1-1 linkage).

For example, consider a 2-1-1 design in which the Level-2 variable *X* is the treatment variable *training on the job*, with one group receiving training and another not receiving training; the Level-1 mediator *M* is *job-related skills*; and the Level-1 outcome variable *Y* is *job performance* (i.e., training influences job-relevant skills, which in turn influences job performance). In this case, *X* varies only between groups (people as a group either do or do not receive training), whereas both *M* and *Y* vary both within and between the groups (people differ from each other within the groups in their skills and performance, and there are differences between the groups in skills and performance). When one estimates the influence of training on job-related skills, training influences individual skills but does so for the group as a whole, making the training effect a between-group effect. Because training was provided to the entire group without differential application across people within the group, it cannot account for within-

group differences of any kind. Again, that is not to say that training has no impact on Level-1 skills; it does, but only because individuals belong to groups that either did or did not receive training.[6] For the same reason, training can impact job performance only at the level of the group. Training cannot account for individual differences within a group in skills and performance (i.e., deviations of individuals from the group mean), because training was applied equally to all members within each group. This logic is similar to that of the ANOVA framework, in which a treatment effect is purely a between-group effect. Therefore, the indirect effect of training (*X*) on job performance (*Y*) through job-relevant skills (*M*) may function only through the between-group variance in *M* and *Y*. It follows that, in a mediation model for 2-1-1 data, when the *b* effect estimate conflates the Within and Between effects, the indirect effect that necessarily operates between groups is confounded with the within-group portion of the conflated *b* effect.

This is an important issue, because multilevel mediation studies have often used MLM without separating the Between and Within components of the 1-1 linkage (for models for 2-1-1 data, see Krull & MacKinnon, 1999, 2001, and Pituch & Stapleton, 2008; for models for 1-1-1 data, see Kenny et al., 2003, and Bauer et al., 2006). In these studies, the Between and Within effects of *M* on *Y* are conflated (i.e., only one mean *b* slope is estimated). However, because each of these studies involved data generated in a way that

———

[6] It is of course possible that individuals may respond differently to training. However, this differential response does not reflect the effect of training (which has no within-cluster variance and thus cannot evoke variability in individual responses within clusters). Rather, it reflects the influence of unmeasured or omitted person-level variables that moderate the influence of cluster-level training.

guaranteed that the Between and Within components were identical (in the data-generating models, only a single conflated effect of M on Y was specified), this led to low bias for the conditions examined. This situation of equal b parameters across levels, we conjecture, is rare in practice. If the Between and Within b effects differ, the single estimated b slope in the conflated model will be a biased estimate of both.

The standard solution for the problem of conflation in MLM has been to disentangle the Within and Between effects of a Level-1 variable by replacing the Level-1 predictor X_{ij} with the two predictors $(X_{ij} - \bar{X}_{.j})$ and $\bar{X}_{.j}$, where $(X_{ij} - \bar{X}_{.j})$ represents the within-group portion of X_{ij} and $\bar{X}_{.j}$ is the group mean of X_{ij} (Hedeker & Gibbons, 2006; Kreft & de Leeuw, 1998; Snijders & Bosker, 1999). We term this approach the *unconflated multilevel model* (UMM) approach, emphasizing that the Between and Within components of the effect of X on Y are no longer conflated into a single estimate. Two prior multilevel mediation studies have used the UMM approach to reduce bias due to conflation (group mean centering M and including both the centered M and the group mean of M as predictors in the equation for Y; MacKinnon, 2008; Zhang, Zyphur, & Preacher, 2009).

However, even though the UMM procedure separates Between and Within effects, a problem remains. Using the group mean as a proxy for a group's latent standing on a given predictor typically results in bias of the between-group effect for the predictor. This effect can be understood by drawing an analogy to latent variable models; Level-1 units can be understood as indicators of Level-2 latent constructs, so in the same way that few indicators and low communalities can bias relationships among latent variables in factor analysis, few Level-1 units and a low intraclass correlation (ICC) can bias Between effects (Asparouhov & Muthén, 2006; Lüdtke et al., 2008; Mancl et al., 2000; Neuhaus & McCulloch, 2006). Lüdtke et al. (2008) showed that the expected bias of the UMM method in estimating the Between effect in a 1-1 design is

$$E(\hat{\gamma}_{Y01} - \beta_B) = (\beta_W - \beta_B) \cdot \frac{1}{n} \cdot \frac{1 - \text{ICC}_X}{\text{ICC}_X + \frac{(1 - \text{ICC}_X)}{n}}, \quad (1)$$

where $\hat{\gamma}_{Y01}$ is the Between effect estimated using observed group means; β_W and β_B are the population Within and Between effects, respectively; n is the constant cluster size; and ICC_X is the intraclass correlation of the predictor X.

The expected bias of the Between indirect effect in a model for 2-1-1 data using UMM (separating M into Between and Within components by group mean centering) is

$$E(\hat{\gamma}_{M01}\hat{\gamma}_{Y01} - \alpha_B\beta_B) = \alpha_B \left(\frac{\beta_B\left(\tau_M^2 - \frac{\tau_{XM}^2}{\tau_X^2}\right) + \beta_W\frac{\sigma_M^2}{n}}{\left(\tau_M^2 - \frac{\tau_{XM}^2}{\tau_X^2}\right) + \frac{\sigma_M^2}{n}} \right) - \alpha_B\beta_B, \quad (2)$$

where $\hat{\gamma}_{M01}$ and $\hat{\gamma}_{Y01}$ are, respectively, the Between effect of X on M and of M on Y estimated with observed group means; α_B is the population Between effect of X on M; τ_X^2, τ_M^2, and τ_{XM} are the Between variances and covariance of X and M; and σ_M^2 is the Within variance of M. The Between effect $\hat{\gamma}_{Y01}$ is a weighted

average of the Between and Within effects of M on Y that depends on the cluster-level variances and covariance of X and M, the within-cluster variance of M, and the cluster size (assumed to be constant over clusters; for full derivation, see Appendix A). As n increases, the Within variance of M carries less weight, yielding an unbiased estimator. In most applications, UMM will yield a biased estimate of the true Between indirect effect. In any mediation model containing a 1-1 component, the Between indirect effect will be similarly biased.

If the Between and Within effects of M on Y in the 2-1-1 design actually do differ, neither the standard MLM approach nor the UMM approach will return unbiased estimates of the Between indirect effect. To mitigate this bias the researcher must use a method that treats the cluster-level component of Level-1 variables as latent. The proposed MSEM framework does this. No prior multilevel mediation studies have used the MSEM approach to return unbiased estimates of the Between indirect effect. We recommend that MSEM be used to specify the model for 2-1-1 data and, indeed, any multilevel mediation hypothesis in which unreliability in groups' standing exists. Using group means as proxies for latent cluster-level means implies that the researcher believes the means are perfectly reliable. To the extent that this is not true, relationships involving at least one such group mean will, on average, be biased relative to the true effect (Asparouhov & Muthén, 2006; Lüdtke et al., 2008; Mancl et al., 2000; Neuhaus & McCulloch, 2006). The MSEM approach not only permits separate (and theoretically unbiased) estimation of the Between and Within components of 1-1 slopes but also includes the conflated MLM approaches of Bauer et al. (2006), Krull and MacKinnon (1999, 2001), and Pituch and Stapleton (2008) as special cases, if this is desired, with the imposition of an equality constraint on the fixed components of the Between and Within 1-1 slopes. Additionally, as with the various MLM methods, the Within slopes may be specified as random effects.

Second Limitation of MLM Framework for Multilevel Mediation: Inability to Treat Upper-Level Variables as Outcomes (Designs 3 to 7 in Table 1)

MLM techniques are also limited in that they cannot accommodate dependent variables measured at Level 2. Indeed, Krull and MacKinnon (2001) indicated that two basic requirements for examining multilevel mediation with MLM are (a) that the outcome variable be assessed at Level 1 and (b) that each effect in the causal chain involves a variable affecting another variable at the same or lower level. In other words, "upward" effects are ruled out by the use of traditional multilevel modeling techniques. Pituch et al. (2006) and Pituch, Murphy, and Tate (2010) also noted that MLM software cannot be used to model 2-2 relationships.

Yet, theoretical predictions of this sort do occur. Effects characterized by a Level-1 variable predicting a Level-2 outcome are known as *bottom-up* effects (Bliese, 2000; Kozlowski & Klein, 2000), *micro-macro* or *emergent* effects (Croon & van Veldhoven, 2007; Snijders & Bosker, 1999), or (*cross-level*) *upward influence* (Griffin, 1997). Bottom-up effects are commonly encountered in practice. For example, family data are intrinsically hierarchical in nature: Parent variables are associated with the family (cluster)

level and children are nested within families, yet a great deal of research shows that children can impact family-level variables (e.g., Bell, 1968; Bell & Belsky, 2008; Kaugars, Klinnert, Robinson, & Ho, 2008; Lugo-Gil & Tamis-LeMonda, 2008). To date, three approaches have been used to circumvent this issue when mediation hypotheses are tested with multilevel data: two-step analyses, aggregation, and disaggregation. Each has limitations, which we discuss in turn.

Two-Step Analyses

Griffin (1997) proposed a two-stage method for analyzing bottom-up effects in which the intercept residuals from a Level-1 equation (estimated with MLM) serve as predictors in a Level-2 equation (estimated with ordinary least squares [OLS] regression). He illustrated the procedure by predicting group-level cohesion using group-level empirical Bayes estimated residuals drawn from the individual-level prediction of affect by concurrent and previous group cohesion. Conversely, Croon and van Veldhoven (2007) applied linear regression analysis predicting a Level-2 outcome from group means of a Level-1 predictor adjusted for the bias that results from using observed means as proxies for latent means. Applying Griffin's (1997) or Croon and van Veldhoven's (2007) methods in the context of mediation analysis would involve multiplying the estimate thus obtained for the 1-2 portion of the design by another coefficient to compute an indirect effect. These methods could, in theory, be used in 1-2-1, 1-2-2, or more elaborate designs. Lüdtke et al. (2008) showed that Croon and van Veldhoven's procedure is characterized by less efficient estimation than a one-stage full-information maximum likelihood procedure. Furthermore, use of these methods would involve an inconvenient degree of manual computation, especially in situations involving more than one predictor and more than one dependent variable (as in mediation models).

For predictors already assessed at Level 2, Krull and MacKinnon (2001) and Pituch et al. (2006) used OLS regression to obtain the *a* component of the indirect effect and MLM to obtain the *b* component for a model for 2-2-1 data (see Takeuchi, Chen, & Lepak, 2009, for an application). Although this method is useful and intuitive, it is more convenient to estimate the parameters making up the indirect effect simultaneously, as part of one model, in which multiple components of variance could be considered simultaneously. The convenience of MSEM relative to two-step analyses becomes more apparent as the models become more complex (e.g., incorporating multiple indirect paths and latent variables with multiple indicators).

Aggregation

Aggregation involves using group means of Level-1 variables in Level-2 equations, such that models for 2-2-1, 1-2-2, 2-1-2, or other multilevel designs are all fit at the cluster level using traditional single-level methods in analysis. Aggregation of data to a higher level is highly problematic, in part because this assumes that the within-group variability of the aggregated variable is zero (Barr, 2008). As is well known, aggregation often (a) leads to severe bias and loss of power for the regression weights composing an indirect effect, (b) discounts information regarding within-

unit variation, (c) drastically reduces the sample size, and (d) increases the risk of committing common fallacies, such as the ecological fallacy and the atomistic fallacy (Alker, 1969; Chou, Bentler, & Pentz, 2000; Diez-Roux, 1998; Firebaugh, 1978; Hofmann, 1997, 2002). Furthermore, aggregation effectively gives small groups and large groups equal weight in determining parameter estimates. Finally, aggregated Level-1 data may not always fairly represent group-level constructs (Blalock, 1979; Chou et al., 2000; Klein & Kozlowski, 2000; Krull & MacKinnon, 2001; van de Vijver & Poortinga, 2002).

Disaggregation

Disaggregation involves ignoring the hierarchical structure altogether and using a single-level model to analyze multilevel mediation relationships. However, disaggregation fails to separate within-group and between-group variance, muddying the interpretation of the indirect effect, and also may spuriously inflate power for the test of the indirect effect (Chou et al., 2000; Julian, 2001; Krull & MacKinnon, 1999).

MSEM

In contrast to these three procedures (two-step, aggregation, and disaggregation), MSEM does not require outcomes to be measured at Level 1, nor does it require two-stage analysis. As we noted earlier, for any mediation model involving at least one Level-2 variable, the indirect effect can exist only at the Between level. So it is not sensible to hypothesize that the effect of a Level-1 X on a Level-2 Y can be mediated. Rather, it is the effect of the *cluster-level component* of X that can be mediated. Within-cluster variation in X necessarily cannot be related to between-cluster variation in M or Y. Nevertheless, it may often be of interest to determine whether or not mediation exists at the cluster level in designs in which a Level-1 variable precedes a Level-2 variable in the causal sequence. Three-variable systems fitting this description include 1-1-2, 1-2-2, 2-1-2, and 1-2-1 designs.

Multilevel SEM

MSEM has received little attention as a framework for multilevel mediation because analytical developments in MSEM have only recently made this approach a feasible alternative to MLM for this purpose. In the past, approaches for fitting MSEMs were hindered by an inability to accommodate random slopes (e.g., Bauer, 2003; Curran, 2003; Härnqvist, 1978; Goldstein, 1987, 1995; McArdle & Hamagami, 1996; Mehta & Neale, 2005; Rovine & Molenaar, 1998, 2000, 2001; Schmidt, 1969) and sometimes also an inability to fully accommodate partially missing data and unbalanced cluster sizes (Goldstein & McDonald, 1988; McDonald, 1993, 1994; McDonald & Goldstein, 1989; Muthén, 1989, 1990, 1991, 1994; Muthén & Satorra, 1995). These early methods typically involved decomposing observed scores into Between and pooled Within covariance matrices and fitting separate Between and Within models using the multiple-group function of many SEM software programs.

dumay.info

On the basis of these and other developments, several authors have specified causal chain models using MSEM, but they did not address indirect effects directly (e.g., Kaplan & Elliott, 1997a, 1997b; Kaplan, Kim, & Kim, 2009; McDonald, 1994; Raykov & Mels, 2007). Although a few authors have implemented multilevel mediation analyses using MSEM for a specific model and have addressed the indirect effect specifically (Bauer, 2003; Heck & Thomas, 2009; Rowe, 2003), they did not allow generalizations for random slopes, did not provide a framework that accommodated all possible multilevel mediation models, and did not provide comparisons to other (MLM) methods (as is done here in the following sections).

These shortcomings of previous work may be attributed largely to technical and practical limitations associated with the two-matrix and other approaches to MSEM. Recent MSEM advances have overcome the limitations of the two-matrix approach regarding random slopes and have also accommodated missing data and unbalanced clusters (Ansari, Jedidi, & Dube 2002; Ansari, Jedidi, & Jagpal, 2000; Chou et al., 2000; Jedidi & Ansari 2001; Muthén & Asparouhov, 2008; Rabe-Hesketh, Skrondal, & Pickles, 2004; Raudenbush, Rowan, & Kang, 1991; Raudenbush & Sampson, 1999; Skrondal & Rabe-Hesketh 2004). However, these methods vary in their benefits, drawbacks, and generality. For example, Raudenbush et al.'s (1991) method involves adding a restrictive measurement model to an otherwise ordinary application of MLM, so it easily allows multiple levels of nesting but does not allow fit indices, unequal factor loadings, or indirect effects at multiple levels simultaneously. These disadvantages are overcome by Chou et al.'s method. However, their method results in still other drawbacks: biased estimation of group-level variability in random coefficients, no correction for unbalanced group sizes, and nonsimultaneous estimation of the Between and Within components of a model. More general methods include those of Ansari and colleagues (Ansari et al., 2000, 2002; Jedidi & Ansari, 2001), Rabe-Hesketh et al. (2004), and Muthén and Asparouhov (2008). Of these, Muthén and Asparouhov's (2008) method is often the most computationally tractable for complex models.[7] In comparison, Ansari et al.'s Bayesian methodology is challenging to implement for the applied researcher; Rabe-Hesketh et al.'s methodology may often require significantly longer computational time (Bauer, 2003; Kamata, Bauer, & Miyazaki, 2008) and does not accommodate random slopes for latent covariates—limiting what can be tested with latent variables and Level-1 predictors and mediators. Therefore, we apply Muthén and Asparouhov's (2008) MSEM approach to the context of mediation in the next section.

A General MSEM Framework for Investigating Multilevel Mediation

The single-level structural equation model can be expressed in terms of a measurement model and a structural model. According to the notation of Muthén and Asparouhov (2008), the measurement model is

$$\mathbf{Y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{X}_i + \boldsymbol{\varepsilon}_i, \qquad (3)$$

where i indexes individual cases; $\mathbf{Y}_i$ is a p-dimensional vector of measured variables; $\boldsymbol{\nu}$ is a p-dimensional vector of variable intercepts; $\boldsymbol{\varepsilon}_i$ is a p-dimensional vector of error terms; $\boldsymbol{\Lambda}$ is a $p \times m$ loading matrix, where m is the number of random effects (latent variables); $\boldsymbol{\eta}_i$ is an $m \times 1$ vector of random effects; and $\mathbf{K}$ is a $p \times q$ matrix of slopes for the q exogenous covariates in $\mathbf{X}_i$. The structural model is

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\zeta}_i, \qquad (4)$$

where $\boldsymbol{\alpha}$ is an $m \times 1$ vector of intercept terms, $\mathbf{B}$ is an $m \times m$ matrix of structural regression parameters, $\boldsymbol{\Gamma}$ is an $m \times q$ matrix of slope parameters for exogenous covariates, and $\boldsymbol{\zeta}_i$ is an m-dimensional vector of latent variable regression residuals. Residuals in $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\zeta}_i$ are assumed to be multivariate normally distributed with zero means and with covariance matrices $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$, respectively.

The model in Equations 3 and 4 expands to accommodate multilevel measurement and structural models by permitting elements of some coefficient matrices to vary at the cluster level. The measurement portion of Muthén and Asparouhov's (2008) general model is expressed as

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j\boldsymbol{\eta}_{ij} + \mathbf{K}_j\mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij} \qquad (5)$$

where j indicates cluster.[8] Note that Equation 5 is identical to Equation 3, but with cluster (j) subscripts on both the variables and the parameter matrices, indicating the potential for some elements of these matrices to vary over clusters. Like $\mathbf{Y}_i$, $\boldsymbol{\nu}$, and $\boldsymbol{\varepsilon}_i$, the matrices $\mathbf{Y}_{ij}$, $\boldsymbol{\nu}_j$, and $\boldsymbol{\varepsilon}_{ij}$ are p-dimensional. Similarly, $\boldsymbol{\Lambda}_j$ is $p \times m$ (where m is the number of latent variables across both Within and Between models and includes random slopes if any are specified), $\boldsymbol{\eta}_{ij}$ is $m \times 1$, and $\mathbf{K}_j$ is $p \times q$. $\mathbf{X}_{ij}$ is a q-dimensional vector of exogenous covariates. The structural component of the general model can be expressed as

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j\boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j\mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}, \qquad (6)$$

where $\boldsymbol{\alpha}_j$ is $m \times 1$, $\mathbf{B}_j$ is $m \times m$, and $\boldsymbol{\Gamma}_j$ is $m \times q$. Residuals in $\boldsymbol{\varepsilon}_{ij}$ and $\boldsymbol{\zeta}_{ij}$ are assumed to be multivariate normally distributed with zero means and with covariance matrices $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$, respectively. Elements in these latter two matrices are not permitted to vary across clusters.

Elements of the parameter matrices $\boldsymbol{\nu}_j$, $\boldsymbol{\Lambda}_j$, $\mathbf{K}_j$, $\boldsymbol{\alpha}_j$, $\mathbf{B}_j$, and $\boldsymbol{\Gamma}_j$ can vary at the Between level. The multilevel part of the general model is expressed in the Level-2 structural model:

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta}_j + \boldsymbol{\gamma}\mathbf{X}_j + \boldsymbol{\zeta}_j \qquad (7)$$

---

[7] The model of Muthén and Asparouhov (2008) is implemented in Mplus. Versions 5 and later of Mplus implement maximum likelihood estimation via an accelerated E-M algorithm described by Lee and Poon (1998) and incorporate further improvements and refinements, some of which are described by Lee and Tsang (1999), Bentler and Liang (2003, 2008), Liang and Bentler (2004), and Asparouhov and Muthén (2007). Unlike earlier methods, the new method can accommodate missing data, unbalanced cluster sizes, and—crucially—random slopes. The default robust maximum likelihood estimator used in Mplus does not require the assumption of normality, yields robust estimates of asymptotic covariances of parameter estimates and χ^2, and is more computationally efficient than previous methods.

[8] We use a notation slightly different from that of Muthén and Asparouhov (2008) to align more closely with standard multilevel modeling notation and to avoid reuse of similar symbols.

It is important to recognize that $\boldsymbol{\eta}_j$ in Equation 7 is very different from the $\boldsymbol{\eta}_{ij}$ in Equations 5 and 6. The vector $\boldsymbol{\eta}_j$ contains all the random effects; that is, it stacks the random elements of all the parameter matrices with j subscripts in Equations 5 and 6—say there are r such random effects. Also, $\mathbf{X}_j$ is different from the earlier $\mathbf{X}_{ij}$. $\mathbf{X}_j$ is an s-dimensional vector of all cluster-level covariates. The vector $\boldsymbol{\mu}$ ($r \times 1$) and matrices $\boldsymbol{\beta}$ ($r \times r$) and $\boldsymbol{\gamma}$ ($r \times s$) contain estimated fixed effects. In particular, $\boldsymbol{\mu}$ contains means of the random effect distributions and intercepts of Between structural equations, $\boldsymbol{\beta}$ contains regression slopes of random effects (i.e., latent variables and random intercepts and slopes) regressed on each other, and $\boldsymbol{\gamma}$ contains regression slopes of random effects regressed on exogenous cluster-level regressors. Cluster-level residuals in $\boldsymbol{\zeta}_j$ have a multivariate normal distribution with zero means and covariance matrix $\boldsymbol{\psi}$.

Together, Equations 5, 6, and 7 define Muthén and Asparouhov's (2008) general model, which we adopt here for specifying and testing multilevel mediation models. In the present context we can safely ignore matrices $\boldsymbol{\nu}_j$, $\mathbf{K}_j$, $\boldsymbol{\Gamma}_j$, and $\boldsymbol{\Theta}$ and we do not require exogenous variables in $\mathbf{X}_{ij}$ or $\mathbf{X}_j$ or residuals $\boldsymbol{\varepsilon}_{ij}$, although all of these are available in the general model. Furthermore, in the models discussed here, $\boldsymbol{\Lambda}_j = \boldsymbol{\Lambda}$. Put succinctly, then,

$$\text{Measurement model: } \mathbf{Y}_{ij} = \boldsymbol{\Lambda}\boldsymbol{\eta}_{ij} \qquad (8)$$

$$\text{Within structural model: } \boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j\boldsymbol{\eta}_{ij} + \boldsymbol{\zeta}_{ij} \qquad (9)$$

$$\text{Between structural model: } \boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta}_j + \boldsymbol{\zeta}_j \qquad (10)$$

with $\boldsymbol{\zeta}_{ij} \sim MVN(\mathbf{0},\boldsymbol{\Psi})$ and $\boldsymbol{\zeta}_j \sim MVN(\mathbf{0},\boldsymbol{\psi})$. Expanding the $\boldsymbol{\eta}_j$ vector clarifies that it contains the random effects, or elements of the matrices $\boldsymbol{\alpha}_j$ and $\mathbf{B}_j$ that potentially vary at the cluster level. Letting the "vec{.}" operator denote a column vector of all the nonzero elements (fixed or random) of one of these matrices, taken in a rowwise manner,

$$\boldsymbol{\eta}_j = \begin{bmatrix} \text{vec}\{\mathbf{B}_j\} \\ \text{vec}\{\boldsymbol{\alpha}_j\} \end{bmatrix} \qquad (11)$$

The vector $\boldsymbol{\mu}$ contains the fixed effects and the means of the random effects:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{B}} \\ \boldsymbol{\mu}_{\boldsymbol{\alpha}} \end{bmatrix} \qquad (12)$$

When all the elements of one of the subvectors of $\boldsymbol{\eta}_j$ have zero means and do not vary at the cluster level, the corresponding subvectors are eliminated from both $\boldsymbol{\eta}_j$ and $\boldsymbol{\mu}$. Within $\boldsymbol{\mu}$, only $\boldsymbol{\mu}_{\mathbf{B}}$ contains parameters involved in the mediation models considered here. Thus, only $\boldsymbol{\mu}_{\mathbf{B}}$ and $\boldsymbol{\beta}$ contain estimated structural parameters, with $\boldsymbol{\mu}_{\mathbf{B}}$ containing the Within effects and $\boldsymbol{\beta}$ containing the Between effects. Throughout, we assume the Within and Between structural models in Equations 9 and 10 to be recursive (i.e., they contain no feedback loops, and both $\mathbf{B}_j$ and $\boldsymbol{\beta}$ can be expressed as lower triangular matrices).

When the general MSEM model is applied to raw data, no explicit centering is required. Rather, the model implicitly partitions each observed Level-1 variable into latent Within and Between components. The MSEM models we discuss here do not involve or require explicit centering of observed predictor variables, but researchers are free to grand mean or group mean center their observed variables as in traditional MLM. Grand mean centering simply rescales a predictor such that the grand mean is subtracted from every score, regardless of cluster membership. Zero corresponds to the grand mean in the centered variable, and other values represent deviations from the grand mean. Group mean centering involves subtracting the cluster mean from every score. Group mean centering thus removes all Between variance in an observed variable prior to modeling, rendering it a strictly Within variable (i.e., all cluster means are equal to zero in the centered variable). On a practical note, the researcher may wish to group mean center a Level-1 predictor variable when its Between variance is essentially zero. If a model is estimated using a variable with a very small ICC, the estimation algorithm may not converge to a proper solution. More generally, the researcher may simply wish to focus on the within-cluster individual differences and have no interest in between-cluster differences and thus may choose to eliminate the Between portion of a predictor in the interests of parsimony. Group mean centering of Level-1 predictors is ill advised in cases where between-cluster effects are of theoretical interest, or are at least plausible in light of theory and potentially of interest; we suspect this accounts for the majority of cases in practice. More is said about centering strategies, and the motivations for centering, by Enders and Tofighi (2007), Hofmann and Gavin (1998), and Kreft et al. (1995).

The general MSEM model in Equations 8–10 contains SEM (and thus factor analysis and path analysis) and two-level MLM as special cases. Although we restrict attention to the single-population, continuous variable MSEM to simplify presentation, Muthén and Asparouhov (2008) presented extensions to this MSEM accommodating individual- and cluster-level latent categorical variables, as well as count, censored, nominal, semicontinuous, and survival data. Future investigations of these extensions in the context of multilevel mediation analysis are certainly warranted. Additional discussion of the general model can be found in Kaplan (2009) and Kaplan et al. (2009).

Specifying Multilevel Mediation Models in MSEM

We next describe how models discussed previously in the multilevel mediation literature, as well as new mediation models for hierarchical data, can be specified as special cases of this general MSEM model. Example Mplus (Muthén & Muthén, 1998–2007) syntax is provided on the first author's website for each of the models discussed here (for Version 5 or later).[9] We begin with the simplest mediation model and follow with exposition of the multilevel mediation model for 2-1-1 data. We address several additional models (for 2-2-1, 1-1-1, 1-1-2, 1-2-2, 2-1-2, and 1-2-1 data) in Appendix B.

Ordinary Single-Level Mediation

It is possible to employ the framework described here when there is no cluster-level variability and only measured variables are used.

_____

[9] See http://www.quantpsy.org

When confronted with such data, most researchers use OLS regression, or more generally path analysis, to assess mediation (Wood, Goodman, Beckmann, & Cook, 2008), following the instructions of numerous methodologists who have investigated mediation analysis (e.g., Baron & Kenny, 1986; Edwards & Lambert, 2007; MacKinnon, 2008; MacKinnon et al., 2002; Preacher & Hayes, 2004, 2008a, 2008b; Shrout & Bolger, 2002). Single-level path analysis is a constrained special case of MSEM with no Between variation, no latent constructs, and no error variance or estimated intercepts for the measured variables. It is not necessary to invoke a multilevel model to assess mediation effects when there is no cluster variation in the variables, but it is informative to see how the general model may be constrained to yield this familiar case by invoking Equations 8, 9, and 10 and letting $\boldsymbol{\beta} = \mathbf{0}, \boldsymbol{\alpha}_j = \mathbf{0}, \mathbf{B}_j = \mathbf{B}, \boldsymbol{\Lambda} = \mathbf{I}$, and $\boldsymbol{\psi} = \mathbf{0}$. We can further grand mean center all variables and let $\boldsymbol{\alpha} = \mathbf{0}$ to render the path analysis model. In the case of single-level path analysis, the general MSEM model reduces to

$$\mathbf{Y}_{ij} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}_{ij} \tag{13}$$

The elements of $\mathbf{B}$ contain the path coefficients making up the indirect effect:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ B_{MXj} & 0 & 0 \\ B_{YXj} & B_{YMj} & 0 \end{bmatrix} \tag{14}$$

Thus,

$$\mathbf{Y}_{ij} = \begin{bmatrix} X_{ij} \\ M_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -B_{MXj} & 1 & 0 \\ -B_{YXj} & -B_{YMj} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \zeta_{Xij} \\ \zeta_{Mij} \\ \zeta_{Yij} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ B_{MXj} & 1 & 0 \\ B_{MXj}B_{YMj} + B_{YXj} & B_{YMj} & 1 \end{bmatrix} \begin{bmatrix} \zeta_{Xij} \\ \zeta_{Mij} \\ \zeta_{Yij} \end{bmatrix}$$

$$\tag{15}$$

The total indirect effect of X on Y through M is given by (Bollen, 1987):

$$\mathbf{F} = (\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I} - \mathbf{B}. \tag{16}$$

Thus,

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ B_{MXj}B_{YMj} & 0 & 0 \end{bmatrix}, \tag{17}$$

in which $B_{MXj}B_{YMj}$ represents the indirect effect of the first column variable (X) on the last row variable (Y) via M. In the more elaborate multilevel models described below and in Appendix B, the same procedure may be used to obtain the indirect effect expressions.

Mediation in 2-1-1 Data

Mediation in 2-1-1 data (also called *upper-level* mediation) has become increasingly common in the multilevel literature, especially in prevention, education, and organizational research (e.g., Kenny, Kashy, & Bolger, 1998; Kenny et al., 2003; Krull & MacKinnon, 2001; MacKinnon, 2008; Pituch & Stapleton, 2008;

Pituch et al., 2006; Tate & Pituch, 2007). For example, Hom et al. (2009) found that job embeddedness (Level 1) mediates the influence of mutual-investment employee–organization relationship (Level 2) on lagged measures of employees' quit intention and affective commitment (Level 1). Similarly, Zhang et al. (2009) hypothesized that job satisfaction (Level 1) mediates the effect of flexible work schedule (Level 2) on employee performance (Level 1). In this example, because flexible work schedule is a cluster-level predictor, it can predict only cluster-level variability in employee performance. Therefore, the question of interest is not simply whether satisfaction mediates the effect of flexible work schedule on performance but whether, and to what degree, cluster-level variability in job satisfaction serves as a mediator of the cluster-level effect of flexible work schedule on the cluster-level component of performance.

The general MSEM model may be constrained to yield a model for 2-1-1 data by invoking Equations 8, 9, and 10. In the case of mediation in 2-1-1 data with random b slopes, the general MSEM equations reduce to

$$\mathbf{Y}_{ij} = \begin{bmatrix} X_j \\ M_{ij} \\ Y_{ij} \end{bmatrix} = \boldsymbol{\Lambda}\boldsymbol{\eta}_{ij} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} \tag{18}$$

$$\boldsymbol{\eta}_{ij} = \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ B_{YMj} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Mij} \\ \zeta_{Yij} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{19}$$

$$\boldsymbol{\eta}_j = \begin{bmatrix} B_{YMj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{BYMj} \\ \mu_{\alpha\eta Xj} \\ \mu_{\alpha\eta Mj} \\ \mu_{\alpha\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta_{MX} & 0 & 0 \\ 0 & \beta_{YX} & \beta_{YM} & 0 \end{bmatrix} \begin{bmatrix} B_{YMj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{BYMj} \\ \zeta_{\alpha\eta Xj} \\ \zeta_{\alpha\eta Mj} \\ \zeta_{\alpha\eta Yj} \end{bmatrix} \tag{20}$$

The partitions serve to separate the Within (above/before the partition) and Between (below/after the partition) parts of the model. Here, η_{Mij} and η_{Yij} are latent variables that vary only within clusters, and η_{Xj}, η_{Mj}, and η_{Yj} vary only between clusters. Because there are only two variables (M and Y) with Within variation, there is no Within indirect effect. The elements of $\boldsymbol{\beta}$ contain the path coefficients making up the Between indirect effect. The total Between indirect effect of X_j on Y_{ij} via M_{ij} are given by extracting the 3×3 Between submatrix of $\boldsymbol{\beta}$ (call it $\boldsymbol{\beta}_B$) and employing Bollen's (1987) formula,

$$\mathbf{F} = (\mathbf{I} - \boldsymbol{\beta}_B)^{-1} - \mathbf{I} - \boldsymbol{\beta}_B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{MX}\beta_{YM} & 0 & 0 \end{bmatrix}, \tag{21}$$

in which $\beta_{MX}\beta_{YM}$ is the Between indirect effect of the first column variable (X_j) on the third row variable (the Between portion of Y_{ij}) via the cluster-level component of M_{ij}. The simultaneous but conflated model of Pituch and Stapleton (2008) may be obtained by constraining $\mu_{BYMj} = \beta_{YM}$. Specifying the Within b slope (B_{YMj})

as fixed rather than random will not alter the formula for the Between indirect effect, although this probably will alter its value and precision somewhat.

Further MSEM Advantages Over MLM for Multilevel Mediation

Of course, the models discussed here and in Appendix B (i.e., for 2-2-1, 2-1-1, 1-1-1, 1-1-2, 1-2-2, 2-1-2, and 1-2-1 data) do not exhaust the two-level possibilities that can be explored with MSEM. Whereas random slopes are permitted to be dependent variables in MLM, the flexibility of MSEM permits examination of models in which random slopes might also (or instead) serve as mediators or independent variables (Cheong, MacKinnon, & Khoo, 2003; Dagne, Brown, & Howe, 2007). Additionally, whereas the addition of multiple mediators (MacKinnon, 2000; Preacher & Hayes, 2008a) becomes a burdensome data management and model specification task in the MLM framework, it is more straightforward in MSEM.

In addition, when MSEM is used, any variable in the system may be specified as a latent variable with multiple measured indicators simply by freeing the relevant loadings. Even these loadings may be permitted to vary across clusters by considering them random effects. In contrast, it is not easy to use latent variables with standard MLM software. In order to incorporate latent variables into MLM, the researcher is obliged to make the strict assumption of equal factor loadings and unique variances (Raudenbush et al., 1991; Raudenbush & Sampson, 1999). Researchers often compute scale means instead. A natural consequence of this practice is that the presence of unreliability in observed variables tends to bias effects downward and thus reduce power for tests of regression parameters. Correction for attenuation due to measurement error is especially important when testing mediation in 1-1-1 data at the within-groups level of analysis, because a majority of the error variance in observed variables exists within groups (because between-group values are aggregates, they will be more reliable; Muthén, 1994).

Finally, within the MLM framework it is not easy to assess model fit, because for many models there is no natural saturated model to which to compare the fit of the estimated model (Curran, 2003), and MLM software and other resources do not emphasize model fit. In the MSEM framework, on the other hand, fit indices often are available to help researchers assess the absolute and relative fit of models. Most of the models discussed here are fully saturated and therefore have perfect fit. However, this will not always be the case in practice, as researchers will frequently wish to assess indirect effects in models with constraints (e.g., measurement models, paths fixed to zero). It will be important to demonstrate that overidentified models have sufficiently good fit before proceeding to the interpretation of indirect effects in such models. A large array of fit indices is available for models without random slopes (Muthén & Muthén, 1998–2007), and information criteria (AIC and BIC) are available for comparing models with random slopes. Lee and Song (2001) demonstrated the use of Bayesian model selection and hypothesis testing in MSEM.[10] Work is under way to develop strategies for assessing model fit in the MSEM context (Ryu, 2008; Ryu & West, 2009; Yuan & Bentler, 2003, 2007).

Empirical Examples

Example 1: Determinants of Math Performance (Mediation in a 1-(1-1)-1 Design)

Our first example is drawn from the Michigan Study of Adolescent Life Transitions (MSALT; Eccles, 1988, 1996). MSALT is a longitudinal study following sixth graders in 12 Michigan school districts over the transition from elementary to junior high school (1983–1985), with two waves of data collection before and two after the transition to seventh grade and several follow-ups. The basic purpose in MSALT is to examine the impact of classroom and family environmental factors as determinants of student development and achievement. This extensive data set combines student questionnaires, academic records, teacher questionnaires and ratings of students, parent questionnaires, and classroom environment measures.

We considered students ($N = 2,993$) nested within teachers ($J = 126$) during Wave 2 of the study (spring of sixth grade). We tested the hypothesis that the effect of student-rated global self-esteem (*gse*) on teacher-rated performance in math (*perform*) was mediated by teacher-rated talent (*talent*) and effort (*effort*). Global self-esteem was computed as a composite of five items assessing students' responses ("sort of true for me" or "really true for me") relative to exemplar descriptions such as "Some kids are pretty sure of themselves" and "Other kids would like to stay pretty much the same." Performance in math was assessed by the single item "Compared to other students in this class, how well is this student performing in math?" ($1 = near the bottom of this class$, $5 = one of the best in the class$). *Talent* and *effort* were assessed with the items "How much natural mathematical talent does this student have?" ($1 = very little$, $7 = a lot$) and "How hard does this student try in math?" ($1 = does not try at all$, $7 = tries very hard$), respectively. ICCs were *gse* (.03), *talent* (.11), *effort* (.13), and *perform* (.04).

Because all four variables were assessed at the student level, this model could be described as a 1-(1,1)-1 model with two mediators with correlated residuals (see Figure 1). We specified random intercepts and fixed slopes, save that the direct effect of *gse* on *perform* was specified as a random slope. Although an MSEM model with all random intercepts and random slopes can be estimated, such a model adds unnecessary complications and may reduce the probability of convergence. The MSEM approach permits us to investigate mediation by multiple mediators simultaneously and at both levels. The Within indirect effect through *talent* (.140) was significant, 90% CI [0.104, 0.178],[11] as was the indirect effect through *effort* (.098), 90% CI [0.073, 0.125]. These

---

[10] There is some ambiguity over whether to use the total sample size, the number of clusters, or the average cluster size in formulas for fit indices in MSEM (Bovaird, 2007; Selig, Card, & Little, 2008). Mehta and Neale (2005) suggested that the number of clusters is the most appropriate sample size to use, whereas Lee and Song (2001) used the total sample size. Total sample size is used in Mplus.

[11] All indirect effect confidence intervals are 90% to correspond to one-tailed, $\alpha = .05$ hypothesis tests, which we feel are often justified in mediation research; contrast CIs are 95% intervals to correspond to two-tailed hypothesis tests.
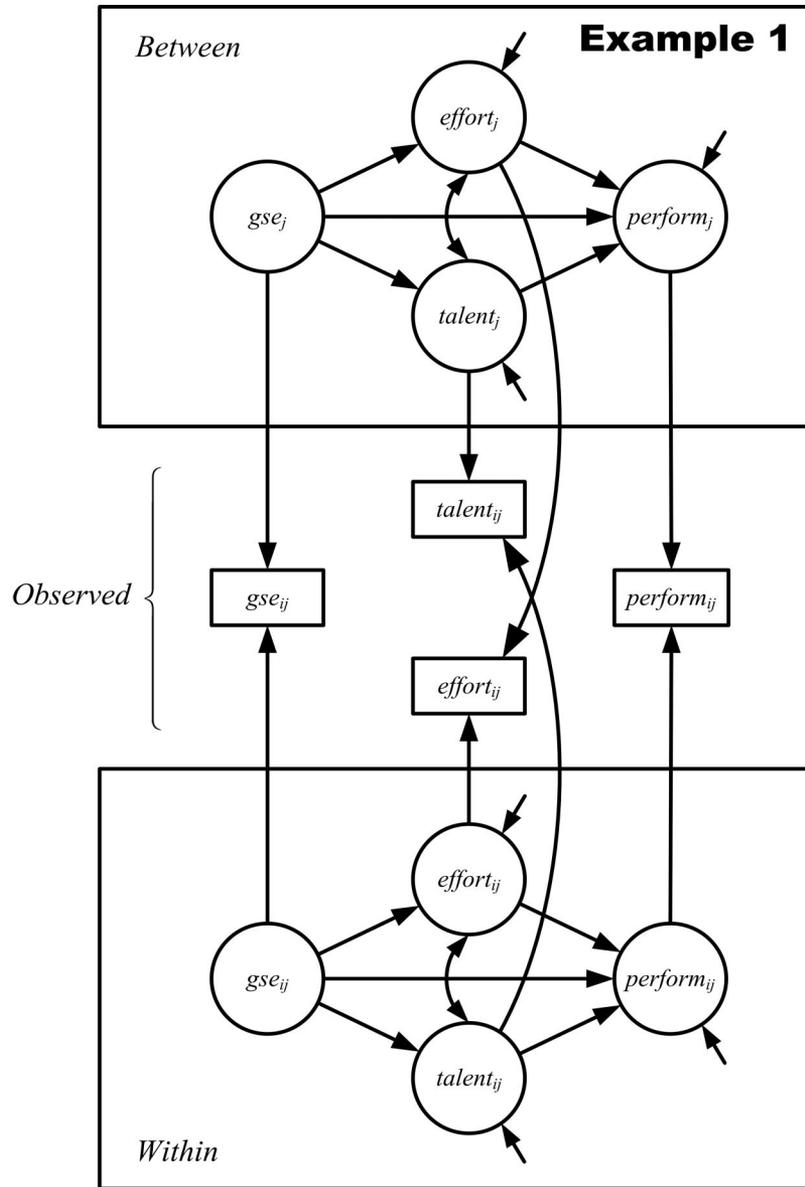
Figure 1. Illustration of a mediation model for 1-1-1 data from Example 1. For simplicity, the random slope of *perform* regressed on *gse* is not depicted. *effort* = teacher-rated effort; *gse* = student-rated global self-esteem; *perform* = teacher-rated performance in math; *talent* = teacher-rated talent.

two indirect effects were not significantly different from each other (contrast = .042), 95% CI [−0.011, 0.096]. The Between indirect effect of *gse* on *perform* via *talent* was significant (.296), 90% CI [0.046, 0.633], as was the indirect effect via *effort* (.182), 90% CI [0.004, 0.430], indicating that teachers whose students tend on average to have higher self-esteem tend to award their students higher ratings for both *talent* and *effort*, and these in turn mediated the effect of group self-esteem on collective math performance. These Between indirect effects were not significantly different (.114), 95% CI [−0.312, 0.589].

The analysis was repeated with the UMM method, that is, by explicitly group mean centering *gse*, *talent*, and *effort* and using the centered variables and their means as predictors of *perform*. As expected, the Within indirect effects through *talent* and *effort*, and the contrast of the two indirect effects, were essentially the same as in MSEM, as were their confidence intervals. The Between indirect effect of *gse* on *perform* via *talent* was significant (.219), 90% CI [0.078, 0.381], as was the indirect effect via *effort* (.130), 90% CI [0.028, 0.250]. This indicates, again, that teachers whose students have higher average self-esteem tend to award higher ratings for both *talent* and *effort*, and these ratings in turn mediate the effect of group self-esteem on collective math performance. These Between indirect effects were not significantly different (.090), 95% CI [−0.132, 0.321]. The noteworthy feature of the

UMM analysis is that, whereas the Within indirect effects are comparable to those from the MSEM analysis, the Between indirect effects are considerably smaller because of bias, although still statistically significant. The Between results are presented in the upper panel of Table 2.

Example 2: Antecedents of Team Potency (Mediation in a 2-2-1 Design)

In our second example, we examine the relationship between a team's shared vision and team members' perceived team potency, as mediated by the team leader's identification with the team. Multi-source survey data were collected from 79 team leaders and 429 team members in a customer service center of a telecommunications company in China. The various scales measuring the variables were translated and back-translated to ensure conceptual consistency. Confidentiality was assured, and the leaders and members returned the completed surveys directly to the researchers. The number of sampled team members ranged from 4 to 7 (excluding the leader), with an average team size of 5.6 members per team.

Leaders reported teams' *shared vision* (three items used in Pearce & Ensley, 2004) and their *identification* with the team (five items used in Smidts, Pruyn, & van Riel, 2001), and members reported their perceived *team potency* (five items used in Guzzo, Yost, Campbell, & Shea, 1993). All three measures were assessed with a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*). Example items include "Team members provide a clear vision of who and what our team is" (*shared vision*), "I experienced a strong sense of belonging to the team" (*identification*), and "Our team feels it can solve any problem it encounters" (*team potency*). We used scale items as indicators for all three latent constructs. The ICCs for team potency indicators ranged from .33 to .48.

Using both the MSEM and conventional two-step approach, we tested the hypothesis that *shared vision* would affect *potency* indirectly via *identification*. Because *shared vision* and *identifica-tion* are leader-level measures and *team potency* is a member-level measure, this is a 2-2-1 design. We specified random intercepts and fixed slopes in both the MSEM and the two-step approaches. Results based on the two approaches are shown in the middle panel of Table 2. Although both approaches showed significant indirect effects, there are differences in the estimates of the path coefficients and the indirect effect itself.

In the two-step approach, the first step involves estimating the 2-2 part of the 2-2-1 design using OLS regression ($J = 79$ team leaders). With leader identification regressed upon leader vision, the unstandardized coefficient (.387, $SE = .094$, $p < .001$) represents path a of the indirect effect. In the second step, leader vision and leader identification both predict team members' perceptions of team potency in a hierarchical linear model with random intercepts and fixed slopes. The coefficient of leader identification (.377, $SE = .149$, $p < .05$) represents path b of the indirect effect. When this two-step approach is used, the indirect effect is $a \times b =$ .146 ($p < .05$), 90% CI [.044, .269]. Note that, although the three variables in this example all have multiple indicators, in this two-step approach the mean of multiple items is typically used to measure the variable of interest when the internal consistency coefficient is satisfactory (e.g., greater than .7).

In the MSEM approach, in contrast, multiple indicators are used to measure the latent variables of leader vision, leader identification, and team members' perceptions of team potency (as shown in Figure 2). MSEM permits us to investigate the (2-2) and (2-1) linkages simultaneously, rather than in two steps as the conventional MLM framework requires. The MSEM estimated indirect effect was higher than that of the two-step approach (.218, $p < .10$, vs. .146, $p < .05$) and had a wider confidence interval, 90% CI [0.055, 0.432]. In addition, the MSEM results showed a larger standard error for the a path than that in the OLS regression of the two-step approach. The MSEM approach provides fit indices for the mediation model (not shown in the table), and these indices showed satisfactory

Table 2

MSEM Versus Conventional Approach Estimates and Standard Errors for Path a, Path b, and the Indirect Effect

Approach	*a* path	*b* path	Indirect effect [90% CI]
Example 1: 1-(1,1)-1 design			
UMM paths via *talent*	0.757** (0.286)	0.290*** (0.050)	.219** [0.078, 0.381]
UMM paths via *effort*	0.659* (0.305)	0.197*** (0.044)	.130* [0.028, 0.250]
MSEM paths via *talent*	1.494* (0.594)	0.198* (0.086)	.296* [0.046, 0.633]
MSEM paths via *effort*	1.190 (0.656)	0.153* (0.065)	.182* [0.004, 0.430]
Example 2: 2-2-1 design			
Two-step approach with manifest variables			
Step 1: OLS for *a* path			
Step 2: MLM for *b* path	0.387*** (0.094)	0.377* (0.149)	.146** [0.044, 0.269]
MSEM with latent variables	0.567** (0.214)	0.385** (0.135)	.218** [0.055, 0.432]
Example 3: 1-1-2 design			
Two-step approach with manifest variables			
Step 1: UMM for *a* path			
Step 2: OLS with aggregated variables for *b* path	1.168*** (0.161)	0.045 (0.171)	.053 [−0.276, 0.387]
MSEM with latent variables	1.714*** (0.331)	0.406* (.239)	.697* [0.022, 1.459]

Note. Standard errors are shown in parentheses. Confidence intervals are based on the Monte Carlo method (available at http://www.quantpsy.org). MSEM = multilevel structural equation modeling; CI = confidence interval; UMM = unconflated multilevel modeling; OLS = ordinary least squares; MLM = traditional multilevel modeling.
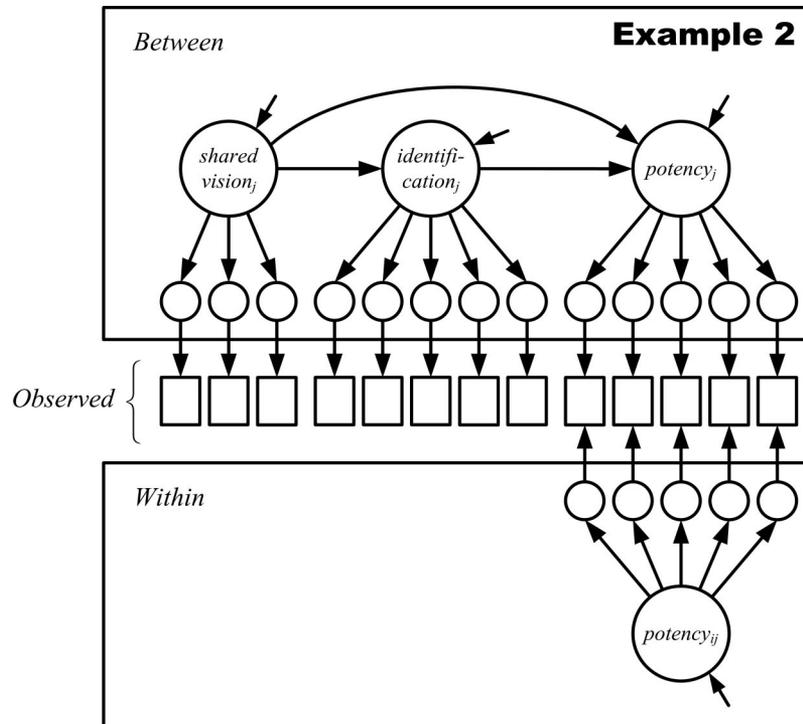* $p < .05$.   ** $p < .01$.   *** $p < .001$ (one-tailed tests).

Figure 2. Illustration of mediation model for 2-2-1 data from Example 2. *shared vision* = leader-reported team shared vision; *identification* = leader-reported member identification with the team; *potency* = member-reported perceived team potency.

model fit (comparative fit index = .94, Tucker–Lewis index = .92, RMSEA = .047, $SRMR_B$ = .08, $SRMR_W$ = .02, where $SRMR_B$ and $SRMR_W$ are the standardized root mean square residuals for the Between and Within models, respectively).

Although the estimates based on the two approaches differ in magnitude, both indicate that teams with a strong shared vision are more likely to have leaders who identify with their teams. This in turn leads to higher team potency as perceived by team members.

Example 3: Leadership and Team Performance (Mediation in a 1-1-2 Design)

Our third example examines an upward influence model in which team members' *perceived transformational leadership* (*TFL*) impacts *team performance*, as mediated by members' *satisfaction with the team*. We utilized the data set from Example 2 to examine this model for 1-1-2 data.

Team members reported their perceptions of transformational leadership (four subdimension scores were based upon 20 items in the Multifactor Leadership Questionnaire; Bass & Avolio, 1995) and satisfaction with the team (three items used in Van der Vegt, Emans, & Van de Vliert, 2000). Higher level managers who are external to the teams' operations rated teams' performance on five items (used in Van der Vegt & Bunderson, 2005). *TFL* was assessed using a 5-point scale (0 = *never*, 4 = *frequently*), and *performance* and *satisfaction* were measured on 7-point Likert scales (1 = *strongly agree*, 7 = *strongly disagree*). Example items include "My leader articulates a compelling vision of the future"

(*TFL*), "I am pleased with the way my colleagues and I work together" (*satisfaction*), and "This team accomplishes tasks quickly and efficiently" (*performance*).

All variables were again treated as latent in the MSEM approach, with items or subdimension scores as indicators. Variables were treated as manifest in the two-step approach. ICCs ranged from .27 to .37. We specified random intercepts and fixed slopes for both the two-step approach and MSEM. Results based on MSEM and the two-step approach are shown in the lower panel of Table 2. The two approaches showed highly different estimates of the indirect effect, which would lead to very different substantive conclusions based on the same data.

In the two-step approach, the first step estimated the 1-1 part of the 1-1-2 design using MLM with random intercepts and fixed slopes. Job satisfaction was regressed upon both the group mean centered *TFL* and the aggregated *TFL* at the team level. The unstandardized coefficient of the aggregated *TFL* (1.168, *SE* = 0.161, *p* < .001) represents the *a* path of the indirect effect. In the second step, we aggregated *TFL* and job satisfaction to the team level and used them to predict leader rated team performance using OLS regression (*J* = 79 teams). The coefficient of aggregated job satisfaction (.045, *SE* = .171, *p* < .80) represents path *b* in the indirect effect. When this two-step approach was used, the indirect effect was not significant, *a* × *b* = .053 (*p* < .85), 90% CI [–0.276, 0.387]. Although the three variables in this example all have multiple indicators, in this two-step approach the means of the multiple items are used to measure the variable of interest when the internal consistency coefficient is satisfactory (e.g., greater than .7).

In the MSEM approach, in contrast, multiple indicators were used to measure the latent variables of *TFL, job satisfaction*, and leader-rated team *performance* (as shown in Figure 3). Paths *a* and *b* were estimated simultaneously rather than in two separate steps. The MSEM estimated indirect effect was significant and much higher than that of the two-step approach ($.697$, $p < .10$, vs. $.053$, $p < .85$). In addition, the MSEM results shows larger estimates for both the *a* path and *b* path. The MSEM approach provides fit indices for the mediation model (not shown in the table), which showed satisfactory model fit (comparative fit index $= .97$, Tucker–Lewis index $= .96$, RMSEA $= .050$, SRMR$_B = .07$, SRMR$_W = .03$).

In sum, the MSEM results show a significant indirect effect of transformational leadership perception on manager-rated team performance via members' satisfaction, $a \times b = .697$, 90% CI [0.022, 1.459]. The MSEM approach allowed us to examine this upward mediation model, whereas the MLM framework that aggregates the member-level measures to the team level failed to provide support for an indirect effect.

Discussion

Summary

Several recent studies have proposed methods for assessing mediation effects using clustered data. These studies have provided methods that target specific kinds of clustered data within the context of MLM. However, these MLM methods now in common use have three limitations critically relevant to mediation analysis. First, these methods frequently assume that Within and Between effects composing the indirect effect are equivalent, an assumption often difficult to justify on substantive grounds (e.g., Bauer et al., 2006; Krull & MacKinnon, 2001; Pituch & Stapleton, 2008), or they separate the Within and Between effects composing the indirect effect in a manner that risks incurring nontrivial bias for the indirect effect (MacKinnon, 2008; Zhang et al., 2009). Second, other MLM methods involve the inconvenient and ad hoc combination of OLS regression and ML estimation intended to model indirect effects with Level-2 *M* and/or *Y* variables (e.g., Pituch et al., 2006). A third limitation that we have not emphasized, but that nevertheless exists, is that multilevel mediation models have, to date, each been presented in isolation and each been geared toward a single design. Therefore, applied researchers have had difficulty establishing an integrated understanding of how these existing models relate to each other and how to adapt them to test new and different hypotheses.

To address these issues, we proposed a general MSEM framework for multilevel mediation that includes all available methods for addressing multilevel mediation as special cases and that can be readily extended to accommodate multilevel mediation models that have not yet been proposed in the literature. Unlike most MSEM approaches described in the literature, the method we use can accommodate random slopes, thus permitting structural coefficients (or function of structural coefficients) to vary randomly across clusters. To demonstrate the flexibility of the MSEM approach, we demonstrated its use with three applied examples. We argue that the general MSEM approach can be used in the overwhelming majority of situations in
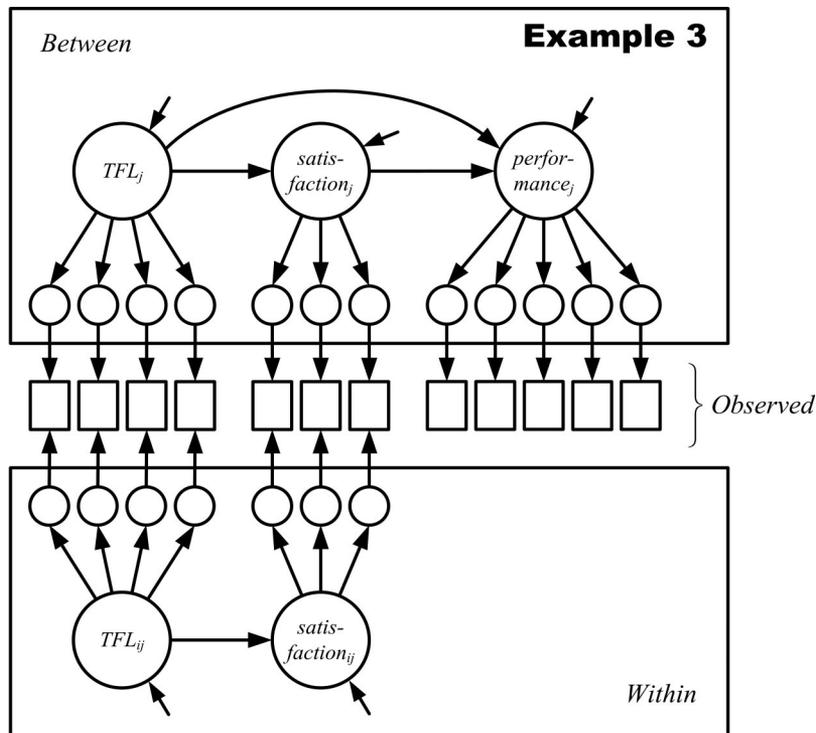

Figure 3. Illustration of mediation model for 1-1-2 data from Example 3. *TFL* = team members' perceived transformational leadership; *satisfaction* = team members' satisfaction with the team; *performance* = manager-rated team performance.

which it is important to address mediation hypotheses with nested data. Below, we address issues relevant to the substantive interpretation of MSEM mediation models, limitations of the approach proposed here, specific concerns related to implementing MSEMs, and a series of recommended steps for estimating multilevel mediation models with MSEM.

Issues Relevant to Interpreting Multilevel Mediation Results

Because the multilevel mediation models described here involve variables and constructs that span two levels of analysis, it is paramount that researchers do more than simply consider the statistical implications of multilevel mediation analyses using MSEM. In order to substantively interpret these models, one should also consider the theoretical implications of variables and constructs used at two levels of analysis via an examination of the linkages among levels of measurement, levels of effect, and the nature of the constructs in question. Such linkages have been explored by many authors (e.g., Bliese, 2000; Chan, 1998; Hofmann, 2002; Klein, Dansereau, & Hall, 1994; Klein & Kozlowski, 2000; Kozlowski & Klein, 2000; Rousseau, 1985; van Mierlo, Vermunt, & Rutte, 2009). These discussions emphasize the importance of establishing the relationships among measurement tools and constructs at multiple levels of analysis, how higher level phenomena emerge over time from the dynamic interaction among lower level parts, as well as how similar constructs are conceptually related at multiple levels of analysis. Although an in-depth exploration of the complexity of issues surrounding multilevel processes, construct validation, and theorizing are beyond the scope of the current paper, it is important to make the following note.

As described above, all multilevel mediation models aside from those for 1-1-1 data describe a mediation effect functioning strictly at the between-group level of analysis. Whether the design suggests top-down effects (e.g., 2-1-1; 2-2-1), bottom-up effects (e.g., 1-1-2; 1-2-2), or both (e.g., 1-2-1; 2-1-2), the mediation effect is inherently at the between-group level of analysis. Because of this fact, researchers must pay close attention to the meaning of lower level variables at the between-group level of analysis and how that meaning is implicated in a particular multilevel mediation effect.

For example, consider individual versus collective efficacy. Whereas individual efficacy originates at the level of the individual, collective efficacy emerges upward across levels of analysis as a function of the dynamic interplay among members of a group (Bandura, 1997). Therefore, simply aggregating individual-level efficacy will not effectively capture the collective efficacy of a group. A self-referential measure of efficacy, such as "I am able to complete my tasks," will assess *individual* efficacy both within and between groups. Although differences between groups on this measure are at the level of the group, the construct is still inherently focused on the individual. In contrast to this, a group-referential measure, such as "my group feels capable of completing its tasks," captures differences within a group in perceptions of the group's collective level of efficacy (which may be thought of as error variance; Chan, 1998), whereas at the between-group level it assesses collective efficacy (Bandura, 1997).

This simple illustration is meant to highlight the potential complexity of multilevel mediation models involving different types of lower level variables. Consider, for example, a mediation model for 2-1-2

data, in which both higher level variables are contextual in nature and the lower level variable is individual efficacy. In this case, latent group standing on individual efficacy mediates the relationship between the higher level variables. Alternatively, consider the same model, where the lower level variable is a measure of collective efficacy. In this case, a fundamentally different form of efficacy—collective efficacy—mediates the relationship. Regardless of the specific form of the multilevel mediation model, researchers should attend to the meaning of a source of variation rather than simply the technical aspects of multilevel mediation point estimates.

Limitations

A limitation of our approach is one shared by most other so-called causal models. Hypotheses of mediation inherently concern conjectures about cause and effect, yet a model is only a collection of assertions that may be shown to be either consistent or inconsistent with hypotheses of cause. Results from a statistical model alone are not grounds for making causal assertions. However, if (a) the researcher is careful to invoke theory to decide upon the ordering of variables in the model, (b) competing explanations for effects are controlled or otherwise eliminated, (c) all relevant variables are considered in the model, (d) data are collected such that conjectured causes precede their conjectured effects in time, and/or (e) X is manipulated rather than observed, claims of causality can be substantially strengthened.

MSEM more generally also has some limitations worth mentioning. The general MSEM presented here cannot employ the restricted maximum likelihood estimation algorithm commonly used in traditional MLM software for obtaining reasonable variance estimates in very small samples. It also cannot easily accommodate designs with more than two hierarchical levels. However, many data sets in education research and social science research are characterized by multiple levels of nesting. Some work has been devoted to mediation in three-level models (Pituch et al., 2010, discussed 3-1-1, 3-2-1, and 3-3-1 designs). The general MSEM could, in theory, be extended to accommodate additional levels (by, for example, permitting the γ and β matrices to vary at Level 3), but the relevant statistical work has not yet been undertaken. This would be a fruitful avenue for future research.

We did not consider models that include moderation of the effects contributing to the Between or Within indirect effects, although such extensions are possible in MSEM. In models involving cross-level interactions, it is possible to find 2-1 effects moderated by Level-1 variables. However, in this case 2-1 main effects still may exist only at the Between level.

Finally, we are careful to limit our discussion to cases with low sampling ratios. That is, we assume two-stage random sampling, in which both Level-2 units and Level-1 units are assumed to have been randomly sampled from infinite populations of such units. If the sampling ratio is high (i.e., when the cluster size is finite and we select a large proportion of individuals from each cluster), the manifest group mean may be a good proxy for group standing (for continuous variables) or group composition (for dichotomous variables such as gender). More discussion of why sampling ratio is important to consider in MSEM is provided by Lüdtke et al. (2008).

Implementation Issues

To scaffold the implementation of the MSEM framework in practice, we now tackle two issues that a researcher would confront when conducting such an analysis: sample size concerns and obtaining confidence intervals for the indirect effect.

Sample size. Questions remain about the appropriate sample size to use for MSEM. To date, no research has investigated the question of appropriate sample size for Muthén and Asparouhov's (2008) model and certainly not in the mediation context. Nevertheless, it is instructive to consult the literature on sample size requirements for precursor MSEM methods for guidance. Investigations of appropriate sample size for MSEM address three issues: the recommended minimum number of units at each level, the degree to which balanced cluster sizes matter, and the diminishing rate of return for collecting additional Level-2 units. Regarding the number of units to collect, Hox and Maas (2001) suggested that at least 100 clusters are required for good performance of Muthén's maximum likelihood (MUML) estimator used with Muthén's two-matrix method, echoing Muthén's (1989) recommendation that at least 50 to 100 clusters should be sampled. Meuleman and Billiet (2009) concluded that as few as 40 clusters may be required to detect large structural paths at the Between level, whereas more than 100 may be necessary to detect small effects (for the models they examined). They also found that having fewer than 20 clusters or a complex Between model can dramatically increase parameter bias. Muthén (1991) and Meuleman and Billiet suggested that, in general, it may be better to observe fewer Level-1 units in favor of observing more Level-2 units. Cheung and Au (2005) suggested that increasing Level-1 sample size also may not help with regard to inadmissible parameter estimates in the Between model. So, on balance, it would seem that emphasizing the number of Level-2 units is particularly important in MSEM, although this question has not been investigated specifically in the context of Muthén and Asparouhov's model, which uses a full ML estimator rather than MUML. Less is known regarding balanced cluster sizes. Hox and Maas (2001) suggested that unbalanced cluster sizes may influence the performance of MSEM using the MUML estimator. On the other hand, Lüdtke et al. (2008) examined unbalanced cluster sizes and found that they did not noticeably affect bias, variability, or coverage when ML was used. Regarding the utility of collecting additional Level-2 units, Yuan and Hayashi (2005) suggested that (when the MUML estimator is used) collecting additional groups may be counterproductive if they contain only a few Level-1 units, although this question, too, has yet to be addressed for Muthén and Asparouhov's method.

Confidence intervals. Our focus has been on model specification rather than on obtaining confidence intervals for the indirect effect. It is possible to use the Mplus MODEL CONSTRAINT command to create functions of model parameters and obtain delta method confidence intervals, but such intervals do not accurately reflect the asymmetric nature of the sampling distribution of an indirect effect. This limitation is problematic, particularly in small samples, but the delta method may be a legitimate approach in large samples. To date, no other procedures have been proposed for obtaining CIs for the indirect effect using MSEM. However, at least three approaches are available for constructing CIs for the indirect effect using MLM—the distribution of the product method, the parametric bootstrap, and the nonparametric boot-

strap—and at least one of them (parametric bootstrap) is generalizable to the MSEM setting.

MacKinnon, Fritz, Williams, and Lockwood (2007); MacKinnon, Lockwood, and Williams (2004); and Williams and MacKinnon (2008) described the *distribution of the product* method for obtaining confidence limits for indirect effects. This method involves using the analytically derived distribution of the product of random variables (in the present case, the product of structural coefficients). This method has the advantage of not assuming symmetry for the sampling distribution of the indirect effect, and therefore confidence intervals based on this method can be appropriately asymmetric. MacKinnon et al. (2007) developed the software program PRODCLIN—now available as a stand-alone executable program and in versions for SAS, SPSS, and R—to enable researchers to use this method.

The *parametric bootstrap* (Efron & Tibshirani, 1986) uses parameter point estimates and their asymptotic variances and covariances to generate random draws from the parameter distributions. In each draw, the indirect effect is computed. This procedure is repeated a very large number of times. The resulting simulation distribution of the indirect effect is used to obtain a percentile CI around the observed indirect effect by identifying the values of the distribution defining the lower and upper $100(\alpha/2)\%$ of simulated statistics. In the single-level mediation context, this method is termed the *Monte Carlo* method (MacKinnon et al., 2004). This method was proposed for use in the MLM context by Bauer et al. (2006); a similar method known as the *empirical-M* method was used in the MLM context by Pituch et al. (2005, 2006) and Pituch and Stapleton (2008). A key feature of the parametric bootstrap is that only the parameter estimates are assumed to be normally distributed. No assumptions are made about the distribution of the indirect effect, which typically is not normally distributed. Pituch et al. (2006) found, in the MLM context, that empirical-M performed nearly as well as the bias-corrected bootstrap, which is quite difficult to implement in multilevel settings. Advantages of this procedure are that it (a) yields asymmetric CIs that are faithful to the skewed sampling distributions of indirect effects, (b) does not require raw data to use, and (c) is very easy to implement. Our empirical examples in this article are the first to implement the parametric bootstrap for the indirect effect in MSEM. To do so, we used Selig and Preacher's (2008) web-based utility to generate and run R code for simulating the sampling distribution of an indirect effect. The utility requires that the user enter values for the point estimates of the *a* and *b* paths, their standard errors (i.e., the square roots of their asymptotic variances), a confidence level (e.g., 90%), and the number of values to simulate.

The *nonparametric bootstrap*, on the other hand, is a family of methods involving resampling the original data (or Level-1 and Level-2 residuals) with replacement, fitting the model a large number of times, and forming a sampling distribution from the estimated indirect effects. A bias-corrected residual-based bootstrap method was implemented in SAS by Pituch et al. (2006) for the conflated model for 2-1-1 data. However, as of this writing, there is no software that can perform both MSEM and the nonparametric bootstrap.

Although only one of the methods proposed for obtaining CIs for the indirect effect in MLM has been used with MSEM (the parametric bootstrap), some methods for obtaining CI for the indirect effect in single-level models could potentially be generalized to the MSEM setting in future research and would be worth

comparing. These include the multivariate delta method (Bollen, 1987; Sobel, 1986) and nonparametric bootstrap methods.

Conclusion

In this article we have advocated a comprehensive MSEM framework for multilevel mediation for the purposes of (a) bias reduction for the indirect effect, (b) unconflating the Between and Within components of the indirect effect, and (c) straightforward generalization to newly proposed multilevel mediation models (Designs 3, 5, 6, and 7 in Table 1). Because of the complexities involved in MSEM, we conclude with the following step-by-step guide (building on Cheung & Au, 2005; Heck, 2001; Muthén, 1989, 1994; Peugh, 2006) for implementing the MSEM framework for multilevel mediation.

1. Identify the mediation hypothesis to be investigated. The first and most critical step is to develop a mediation model that is consistent with theory and includes variables at the appropriate levels of analysis. If there is at least one Level-2 variable in the causal chain, the indirect effect of interest likely is a Between indirect effect. If all three variables are assessed at Level 1, it becomes possible to entertain mediation hypotheses at both levels. As a caution, when researchers lack sufficient theory to posit an appropriate Between model (Peugh, 2006), they often end up specifying identical Within and Between models without justification. This is problematic because the Between model almost always will be different than the Within model (Cronbach, 1976; Cronbach & Webb, 1975; Grice, 1966; Härnqvist, 1978; Kenny & La Voie, 1985; Sirotnik, 1980).

2. Ensure that there is sufficient between-cluster variability to support MSEM. If more than a couple of the variables to be modeled have ICCs below .05, it is likely that convergence will be a problem or that estimation of the indirect effect will be unstable with potentially large bias. Should this happen, the researcher should consider group mean centering the offending variables for the sake of stability, as multilevel techniques may not be profitable or even possible otherwise.

3. Fit the Within model. A properly specified within-cluster model is necessary, but not sufficient, for a properly specified between-cluster model (Heck, 2001; Hofmann, 2002; Hox & Maas, 2001; Kozlowski & Klein, 2000; Muthén & Satorra, 1995; Peugh, 2006). Therefore, we recommend that the Within model be investigated first. There are a few ways to do this. A somewhat crude method involves group mean centering all observed variables and fitting the Within model as a single-level model. Alternatively, the researcher might fit the full model, permitting the cluster-level constructs to covary freely. If the latter approach is taken, it is sensible to begin with few random effects and build in complexity until all theoretically motivated random effects have been included.

4. Fit the Within and Between models simultaneously. Given a well-fitting and interpretable Within model, the researcher might proceed to fitting the full theorized model. At this stage, structure may be imposed on the Between-level constructs and the Between indirect effect may be estimated. Parameter estimates from simpler models may be useful as starting values for estimation in more complex models.

References

Alker, H. R. (1969). A typology of ecological fallacies. In M. Dogan & S. Rokkam (Eds.), *Social ecology* (pp. 69–86). Boston, MA: MIT Press.

Ansari, A., Jedidi, K., & Dube, L. (2002). Heterogeneous factor analysis models: A Bayesian approach. *Psychometrika, 67,* 49–78.

Ansari, A., Jedidi, K., & Jagpal, S. (2000). A hierarchical Bayesian approach for modeling heterogeneity in structural equation models. *Marketing Science, 19,* 328–347.

Asparouhov, T., & Muthén, B. (2006). *Constructing covariates in multilevel regression* (Mplus Web Notes No. 11). Retrieved from http://www.statmodel.com

Asparouhov, T., & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. *Proceedings of the 2007 Joint Statistical Meetings, Section on Statistics in Epidemiology* (pp. 2531–2535). Alexandria, VA: American Statistical Association.

Bacharach, S. B., Bamberger, P. A., & Doveh, E. (2008). Firefighters, critical incidents, and drinking to cope: The adequacy of unit-level performance resources as a source of vulnerability and protection. *Journal of Applied Psychology, 93,* 155–169.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: Worth.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Barr, C. D. (2008, April). *A multilevel modeling alternative to aggregating data in 360 degree feedback.* Poster session presented at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Bass, B. M., & Avolio, B. J. (1995). *Multifactor Leadership Questionnaire technical report.* Palo Alto, CA: Mind Garden.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28,* 135–167.

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.

Bell, B. G., & Belsky, J. (2008). Parents, parenting, and children's sleep problems: Exploring reciprocal effects. *British Journal of Developmental Psychology, 26,* 579–593.

Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological Review, 75,* 81–95.

Bentler, P. M., & Liang, J. (2003). Two-level mean and covariance structures: Maximum likelihood via an EM algorithm. In S. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 53–70). Mahwah, NJ: Erlbaum.

Bentler, P. M., & Liang, J. (2008). A unified approach to two-level structural equation models and linear mixed effects models. In D. B. Dunson (Ed.), *Random effect and latent variable model selection* (pp. 95–119). New York, NY: Springer.

Blalock, H. M. (1979). The presidential address: Measurement and conceptualization problems: The major obstacle to integrating theory and research. *American Sociological Review, 44,* 881–894.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology, 17,* 37–69.

Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors.

In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 149–182). Mahwah, NJ: Erlbaum.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83,* 234–246.

Cheong, J., MacKinnon, D. P., & Khoo, S.-T. (2003). Investigation of mediational processes using parallel process latent growth curve modeling. *Structural Equation Modeling, 10,* 238–262.

Cheung, M. W. L., & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling, 12,* 598–619.

Chou, C., Bentler, P. M., & Pentz, M. A. (2000). A two-stage approach to multilevel structural equation models: Application to longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples* (pp. 33–50). Mahwah, NJ: Erlbaum.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis.* Stanford, CA: Stanford University Evaluation Consortium.

Cronbach, L. J., & Webb, N. (1975). Between-class and within-class effects of repeated aptitude $\times$ treatment interaction: Re-analysis of a study by G. L. Anderson. *Journal of Educational Psychology, 67,* 717–724.

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

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38,* 529–569.

Dagne, G. A., Brown, C. H., & Howe, G. W. (2007). Hierarchical modeling of sequential behavioral data: Examining complex association patterns in mediation models. *Psychological Methods, 12,* 298–316.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analyses. *American Journal of Public Health, 88,* 216–222.

Eccles, J. S. (1988). *Achievement beliefs and the environment* (Report No. NICHD-PS-019736). Bethesda, MD: National Institute of Child Health and Development.

Eccles, J. S. (1996). *Study of Life Transitions (MSALT), 1983–1985.* Retrieved from Henry A. Murray Research Archive at Harvard University website: hdl:1902.1/01015UNF:3:8V65WzuZplrx+MYO8PtzHQ== Murray Research Archive [Distributor].

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12,* 1–22.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science, 1,* 54–75.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121–138.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review, 43,* 557–572.

Goldstein, H. (1987). *Multilevel models in educational and social research.* London, England: Griffin.

Goldstein, H. (1995). *Multilevel statistical models.* New York, NY: Halsted.

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika, 53,* 455–467.

Grice, G. R. (1966). Dependence of empirical laws upon the source of experimental variation. *Psychological Bulletin, 66,* 488–498.

Griffin, M. A. (1997). Interaction between individuals and situations: Using HLM procedures to estimate reciprocal relationships. *Journal of Management, 23,* 759–773.

Guzzo, R. A., Yost, P. R., Campbell, R. J., & Shea, G. P. (1993). Potency in groups: Articulating a construct. *British Journal of Social Psychology, 32,* 87–106.

Härnqvist, K. (1978). Primary mental abilities of collective and individual levels. *Journal of Educational Psychology, 70,* 706–716.

Heck, R. H. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 89–128). Mahwah, NJ: Erlbaum.

Heck, R. H., & Thomas, S. L. (2009). *An introduction to multilevel modeling techniques* (2nd ed.). New York, NY: Routledge.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis.* Hoboken, NJ: Wiley.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23,* 723–744.

Hofmann, D. A. (2002). Issues in multilevel research: Theory development, measurement, and analysis. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 247–274). Malden, MA: Blackwell.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24,* 623–641.

Hom, P. W., Tsui, A. S., Wu, J. B., Lee, T. W., Zhang, A. Y., Fu, P. P., & Li, L. (2009). Explaining employment relationships with social exchange and job embeddedness. *Journal of Applied Psychology, 94,* 277–297.

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

Jedidi, K., & Ansari, A. (2001). Bayesian structural equation models for multilevel data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 129–157). New York, NY: Erlbaum.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling, 8,* 325–352.

Kamata, A., Bauer, D. J., & Miyazaki, Y. (2008). Multilevel measurement models. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 345–388). Charlotte, NC: Information Age.

Kaplan, D. (2009). *Structural equation modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Kaplan, D., & Elliott, P. R. (1997a). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling, 4,* 1–24.

Kaplan, D., & Elliott, P. R. (1997b). A model-based approach to validating education indicators using multilevel structural equation modeling. *Journal of Educational and Behavioral Statistics, 22,* 323–347.

Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. Millsap & A. Maydeu-Olivares (Eds.), *Sage handbook of quantitative methods in psychology* (pp. 592–612). Thousand Oaks, CA: Sage.

Kaugars, A. S., Klinnert, M. D., Robinson, J., & Ho, M. (2008). Reciprocal influences in children's and families' adaptation to early childhood wheezing. *Health Psychology, 27,* 258–267.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 233–265). Boston, MA: McGraw-Hill.

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

Kenny, D. A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology, 48,* 339–348.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology, 83,* 126–137.

Klein, K. J., Conn, A. B., Smith, D. B., & Sorra, J. S. (2001). Is everyone in agreement? An exploration of within-group agreement in employee

perceptions of the work environment. *Journal of Applied Psychology, 86,* 3–16.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review, 19,* 195–229.

Klein, K. J., & Kozlowski, S. W. J. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods, 3,* 211–236.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling.* Thousand Oaks, CA: Sage.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effects of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30,* 1–21.

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

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

Lee, S.-Y., & Poon, W. Y. (1998). Analysis of two-level structural equation models via EM type algorithms. *Statistica Sinica, 8,* 749–766.

Lee, S.-Y., & Song, X.-Y. (2001). Hypothesis testing and model comparison in two-level structural equation models. *Multivariate Behavioral Research, 36,* 639–655.

Lee, S.-Y., & Tsang, S. Y. (1999). Constrained maximum likelihood estimation of two-level covariance structure model via EM type algorithms. *Psychometrika, 64,* 435–450.

Liang, J., & Bentler, P. M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika, 69,* 101–122.

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.

Lugo-Gil, J., & Tamis-LeMonda, C. S. (2008). Family resources and parenting quality: Links to children's cognitive development across the first 3 years. *Child Development, 79,* 1065–1085.

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

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* New York, NY: Erlbaum.

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

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–104.

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

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

Mancl, L., Leroux, B., & DeRouen, T. (2000). Between-subject and within-subject statistical information in dental research. *Journal of Dental Research, 79,* 1778–1781.

Mathieu, J. E., & Taylor, S. R. (2007). A framework for testing meso-mediational relationships in organizational behavior. *Journal of Organizational Behavior, 28,* 141–172.

Maynard, M. T., Mathieu, J. E., Marsh, W. M., & Ruddy, T. M. (2007). A multilevel investigation of the influences of employees' resistance to empowerment. *Human Performance, 20,* 147–171.

McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 89–124). Mahwah, NJ: Erlbaum.

McDonald, R. P. (1993). A general model for two-level data with responses missing at random. *Psychometrika, 58,* 575–585.

McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods and Research, 22,* 399–413.

McDonald, R. P., & Goldstein, H. I. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology, 42,* 215–232.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equation modeling. *Psychological Methods, 10,* 259–284.

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods, 3,* 45–58.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54,* 557–585.

Muthén, B. O. (1990). *Mean and covariance structure analysis of hierarchical data* (UCLA Statistics Series No. 62). Los Angeles: University of California.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28,* 338–354.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22,* 376–398.

Muthén, B. O., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology 1997, 25,* 267–316.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide: Statistical analysis with latent variables* (5th ed.). Los Angeles, CA: Author.

Neuhaus, J., & Kalbfleisch, J. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics, 54,* 638–645.

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

Pearce, C. L., & Ensley, M. D. (2004). A reciprocal and longitudinal investigation of the innovation process: The central role of shared vision in product and process innovation teams (PPITs). *Journal of Organizational Behavior, 25,* 259–278.

Peugh, J. L. (2006). *Specification searches in multilevel structural equation modeling: A Monte Carlo investigation* (Unpublished doctoral dissertation). University of Nebraska.

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

Pituch, K. A., & Stapleton, L. M. (2008). The performance of methods to test upper-level mediation in the presence of nonnormal data. *Multivariate Behavioral Research, 43,* 237–267.

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

Pituch, K. 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.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for

estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36,* 717–731.

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

Preacher, K. J., & Hayes, A. F. (2008b). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks, CA: Sage.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69,* 167–190.

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

Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model of studying school climate with estimation via the EM algorithm and application to U.S. high school data. *Journal of Educational Statistics, 16,* 296–330.

Raudenbush, S. W., & Sampson, R. (1999). Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research, 28,* 123–153.

Raykov, T., & Mels, G. (2007). Lower level mediation effect analysis in two-level studies: A note on a multilevel structural equation modeling approach. *Structural Equation Modeling, 14,* 636–648.

Rousseau, D. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. In L. Cummings & B. Saw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 1–37). Greenwich, CT: JAI.

Rovine, M. J., & Molenaar, P. C. M. (1998). A nonstandard method for estimating a linear growth model in LISREL. *International Journal of Behavioral Development, 22,* 453–473.

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35,* 51–88.

Rovine, M. J., & Molenaar, P. C. M. (2001). A structural equations modeling approach to the general linear mixed model. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 65–96). Washington, DC: American Psychological Association.

Rowe, K. J. (2003). Estimating interdependent effects among multilevel composite variables in psychosocial research: An example of the application of multilevel structural equation modeling. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 255–283). Mahwah, NJ: Erlbaum.

Ryu, E. (2008). *Evaluation of model fit in multilevel structural equation modeling: Level-specific model fit evaluation and the robustness to nonnormality* (Unpublished dissertation). Arizona State University, Tempe.

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling, 16,* 583–601.

Schmidt, W. H. (1969). *Covariance structure analysis of the multivariate random effects model* (Unpublished doctoral dissertation). University of Chicago.

Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modeling in cross-cultural research: Multigroup and multilevel approaches. In F. J. R. van de Vijver, D. A. van Hemert, & Y. Poortinga (Eds.), *Individuals and cultures in multi-level analysis.* Mahwah, NJ: Erlbaum.

Selig, J. P., & Preacher, K. J. (2008, June). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Computer software]. Retrieved from http://www.quantpsy.org

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

Sirotnik, K. A. (1980). Psychometric implications of the unit-of-analysis problem (with examples from the measurement of organizational climate). *Journal of Educational Measurement, 17,* 245–282.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models.* Boca Raton, FL: Chapman & Hall/CRC.

Smidts, A. H., Pruyn, A. T., & van Riel, C. B. M. (2001). The impact of employee communication and perceived external prestige on organizational identification. *Academy of Management Journal, 49,* 1051–1062.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage.

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

Takeuchi, R., Chen, G., & Lepak, D. P. (2009). Through the looking glass of a social system: Cross-level effects of high-performance work systems on employees' attitudes. *Personnel Psychology, 62,* 1–29.

Tate, R. L., & Pituch, K. A. (2007). Multivariate hierarchical linear modeling in randomized field experiments. *Journal of Experimental Education, 75,* 317–337.

Van der Vegt, G. S., & Bunderson, J. S. (2005). Learning and performance in multidisciplinary teams: The importance of collective team identification. *Academy of Management Journal, 48,* 532–547.

Van der Vegt, G. S., Emans, B. J. M., & Van de Vliert, E. (2000). Team members' affective responses to patterns of intragroup interdependence and job complexity. *Journal of Management, 26,* 633–655.

van de Vijver, F. J. R., & Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology, 33,* 141–156.

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing group-level constructs from individual-level survey data. *Organizational Research Methods, 12,* 368–392.

Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling, 15,* 23–51.

Wood, R. E., Goodman, J. S., Beckmann, N., & Cook, A. (2008). Mediation testing in management research: A review and proposals. *Organizational Research Methods, 11,* 270–295.

Yuan, K.-H., & Bentler, P. M. (2003). Eight test statistics for multilevel structural equation models. *Computational Statistics & Data Analysis, 44,* 89–107.

Yuan, K.-H., & Bentler, P. M. (2007). Multilevel covariance structure analysis by fitting multiple single-level models. *Sociological Methodology, 37,* 53–82.

Yuan, K.-H., & Hayashi, K. (2005). On Muthén's maximum likelihood for two-level covariance structure models. *Psychometrika, 70,* 147–167.

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.

(*Appendices follow*)

Appendix A

Bias Derivation

In this Appendix we formally derive the expected bias that results from computing a Between indirect effect in models for 2-1-1 designs using group mean centering for M and entering the group mean of M ($\overline{M}_{.j}$) as an additional predictor of Y (what we term the UMM approach). The developments in this Appendix parallel those in the Appendix of Lüdtke et al. (2008).

The following population models will be assumed for X, M, and Y:

$$X_j = \mu_X + U_{Xj}$$

$$M_{ij} = \mu_M + U_{Mj} + R_{Mij}$$

$$Y_{ij} = \mu_Y + U_{Yj} + R_{Yij}, \tag{A1}$$

where the μs are population means, the U terms are deviations from the means for cluster j, and the R terms are individual deviations from the cluster means. Because X is a cluster-level variable in the 2-1-1 design, there is no within-cluster deviation R_{Xij}. We assume the Us and Rs to be independent. The Within and Between covariance matrices for X, M, and Y are

$$\Sigma_W = \begin{bmatrix} 0 & & \\ 0 & \sigma_M^2 & \\ 0 & \sigma_{MY} & \sigma_Y^2 \end{bmatrix} \tag{A2}$$

$$\Sigma_B = \begin{bmatrix} \tau_X^2 & & \\ \tau_{XM} & \tau_M^2 & \\ \tau_{XY} & \tau_{MY} & \tau_Y^2 \end{bmatrix}. \tag{A3}$$

We are interested in estimating the parameters of the following model,

$$M_{ij} = \mu_M + \alpha_B U_{Xj} + \delta_{Mj} + \varepsilon_{Mij}$$

$$Y_{ij} = \mu_Y + \beta_W R_{Mij} + \beta_B U_{Mj} + \chi_B' U_{Xj} + \delta_{Yj} + \varepsilon_{Yij}, \tag{A4}$$

where μ_M and μ_Y are the conditional means of M and Y, α_B is the Between effect of X on M, β_B is the Between effect of M on Y controlling for X, χ_B' is the Between effect of X on Y controlling for M, β_W is the Within effect of M on Y, δ_{Mj} and δ_{Yj} are cluster-level residuals, and ε_{Mij} and ε_{Yij} are individual-level residuals. Following the UMM strategy, the following model is used to estimate the parameters of A4,

$$M_{ij} = \gamma_{M00} + \gamma_{M01} X_j + u_{Mj} + e_{Mij}$$

$$Y_{ij} = \gamma_{Y00} + \gamma_{Y10}(M_{ij} - \overline{M}_{.j}) + \gamma_{Y01}\overline{M}_{.j} + \gamma_{Y02} X_j + u_{Yj} + e_{Yij}, \tag{A5}$$

where $\overline{M}_{.j}$ is the cluster mean of M, $\hat{\gamma}_{M01}$ is the estimator for α_B, $\hat{\gamma}_{Y01}$ estimates β_B, $\hat{\gamma}_{Y02}$ estimates χ_B', and $\hat{\gamma}_{Y10}$ estimates β_W.

Assuming equal cluster sizes n, the observed covariance matrix $\mathbf{S}$ of Y_{ij}, $(M_{ij} - \overline{M}_{.j})$, $\overline{M}_{.j}$, and X_j can be represented in terms of A2 and A3 as

(Appendices continue)

$$\mathbf{S} = \operatorname{cov}\begin{bmatrix} Y_{ij} \\ M_{ij} - \overline{M}_{\cdot j} \\ \overline{M}_{\cdot j} \\ X_j \end{bmatrix} = \begin{bmatrix} \sigma_Y^2 + \tau_Y^2 & & & \\ \sigma_{MY}\left(1 - \dfrac{1}{n}\right) & \sigma_M^2\left(1 - \dfrac{1}{n}\right) & & \\ \tau_{MY} + \dfrac{\sigma_{MY}}{n} & 0 & \tau_M^2 + \dfrac{\sigma_M^2}{n} & \\ \tau_{XY} & 0 & \tau_{XM} & \tau_X^2 \end{bmatrix}. \tag{A6}$$

If we make use of OLS formulas for regression weights in two- and three-variable systems, the Between effect of X on Y can be derived in terms of A6 as follows:

$$\gamma_{M01} = \frac{\operatorname{cov}(\overline{M}_{\cdot j}, X_j)}{\operatorname{var}(X_j)} = \frac{\tau_{XM}}{\tau_X^2} = \alpha_B. \tag{A7}$$

Thus, $\hat{\gamma}_{M01}$ is an unbiased estimate of α_B. The Between effect of M on Y controlling for X can be derived as

$$\gamma_{Y01} = \frac{r_{Y\overline{M}} - r_{YX} r_{\overline{M}X}}{1 - r_{\overline{M}X}^2} \frac{s_Y}{s_{\overline{M}}} = \frac{\dfrac{\tau_{MY} + \dfrac{\sigma_{MY}}{n}}{\sqrt{\sigma_Y^2 + \tau_Y^2}\sqrt{\tau_M^2 + \dfrac{\sigma_M^2}{n}}} - \dfrac{\tau_{XY}}{\sqrt{\sigma_Y^2 + \tau_Y^2}\sqrt{\tau_X^2}} \dfrac{\tau_{XM}}{\sqrt{\tau_M^2 + \dfrac{\sigma_M^2}{n}}\sqrt{\tau_X^2}}}{1 - \dfrac{\tau_{XM}^2}{\left(\tau_M^2 + \dfrac{\sigma_M^2}{n}\right)\tau_X^2}} \frac{\sqrt{\sigma_Y^2 + \tau_Y^2}}{\sqrt{\tau_M^2 + \dfrac{\sigma_M^2}{n}}}$$

$$= \frac{\dfrac{\operatorname{cov}(Y_{ij}, \overline{M}_{\cdot j})}{\sqrt{\operatorname{var}(Y_{ij})}\sqrt{\operatorname{var}(\overline{M}_{\cdot j})}} - \dfrac{\operatorname{cov}(Y_{ij}, X_j)}{\sqrt{\operatorname{var}(Y_{ij})}\sqrt{\operatorname{var}(X_j)}}\dfrac{\operatorname{cov}(\overline{M}_{\cdot j}, X_j)}{\sqrt{\operatorname{var}(\overline{M}_{\cdot j})}\sqrt{\operatorname{var}(X_j)}}}{1 - \dfrac{\operatorname{cov}^2(\overline{M}_{\cdot j}, X_j)}{\operatorname{var}(\overline{M}_{\cdot j})\operatorname{var}(X_j)}} \frac{\sqrt{\operatorname{var}(Y_{ij})}}{\sqrt{\operatorname{var}(\overline{M}_{\cdot j})}}$$

$$= \frac{\tau_{MY} - \dfrac{\tau_{XY}\tau_{XM}}{\tau_X^2} + \dfrac{\sigma_{MY}}{n}}{\tau_M^2 - \dfrac{\tau_{XM}^2}{\tau_X^2} + \dfrac{\sigma_M^2}{n}} = \frac{\left(\tau_{MY} - \dfrac{\tau_{XY}\tau_{XM}}{\tau_X^2}\right)\dfrac{\sqrt{\tau_Y^2}\sqrt{\tau_M^2}}{\sqrt{\tau_Y^2}\sqrt{\tau_M^2}} + \dfrac{\sigma_{MY}}{n}}{\left(\tau_M^2 - \dfrac{\tau_{XM}^2}{\tau_X^2}\right) + \dfrac{\sigma_M^2}{n}}$$

$$= \frac{\beta_B\left(\tau_M^2 - \dfrac{\tau_{XM}^2}{\tau_X^2}\right) + \beta_W\dfrac{\sigma_M^2}{n}}{\left(\tau_M^2 - \dfrac{\tau_{XM}^2}{\tau_X^2}\right) + \dfrac{\sigma_M^2}{n}}, \tag{A8}$$

which can be understood as a weighted average of β_B and β_W that depends on the cluster-level variances and covariances of X and M, the within-cluster variance of M, and the constant cluster size. When we use the results in A7 and A8, the bias in the Between indirect effect is

$$E(\hat{\gamma}_{M01}\hat{\gamma}_{Y01} - \alpha_B\beta_B) = \alpha_B\left(\frac{\beta_B\left(\tau_M^2 - \dfrac{\tau_{XM}^2}{\tau_X^2}\right) + \beta_W\dfrac{\sigma_M^2}{n}}{\left(\tau_M^2 - \dfrac{\tau_{XM}^2}{\tau_X^2}\right) + \dfrac{\sigma_M^2}{n}}\right) - \alpha_B\beta_B. \tag{A9}$$

As $n \to \infty$, the weight associated with the within-cluster variance of M goes to zero, leaving an unbiased estimator.

Appendix B

Specifying Additional Multilevel Mediation Models in MSEM

We addressed how the general MSEM approach may be applied in the case of single-level data and 2-1-1 data. Here we provide additional guidance for how the general MSEM approach may be used to address mediation in 2-2-1, 1-1-1, 1-1-2, 1-2-2, 2-1-2, and 1-2-1 data. Mplus syntax for all of the models discussed in this paper is provided at the first author's website (http://www.quantpsy.org).

Mediation in 2-2-1 Designs

In the first model we address, a cluster-level mediator is conjectured to mediate the effect of a cluster-level X_j on a Level-1 Y_{ij} (e.g., Takeuchi et al., 2009; for other treatments, see Bauer, 2003, and Heck & Thomas, 2009). In previous literature concerning mediation in 2-2-1 designs it has seemingly gone unrecognized that X_j and M_j can affect only the cluster-level variation in the Level-1 variable Y_{ij}. That is, the only mediation that can occur in the 2-2-1 design occurs at the cluster level. It is the researcher's task to identify whether and to what extent M_j explains the effect of X_j on group standing on Y_{ij}.

There are no 1-1 linkages in the model for 2-2-1 data, so no slopes are random. The intercept of the Y_{ij} equation, however, can be random (Krull & MacKinnon, 2001; Pituch et al., 2006). The general model may be constrained to yield the model for 2-2-1 data, reducing to the equations

$$\mathbf{Y}_{ij} = \begin{bmatrix} X_j \\ M_j \\ Y_{ij} \end{bmatrix} = \mathbf{\Lambda}\mathbf{\eta}_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} \tag{B1}$$

$$\mathbf{\eta}_{ij} = \begin{bmatrix} \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Yij} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{B2}$$

$$\mathbf{\eta}_j = \begin{bmatrix} \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{\alpha \eta Xj} \\ \mu_{\alpha \eta Mj} \\ \mu_{\alpha \eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ \beta_{MX} & 0 & 0 \\ \beta_{YX} & \beta_{YM} & 0 \end{bmatrix} \begin{bmatrix} \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha \eta Xj} \\ \zeta_{\alpha \eta Mj} \\ \zeta_{\alpha \eta Yj} \end{bmatrix}. \tag{B3}$$

Because there is only one variable (Y) with Within variation, there is no Within indirect effect. The $\mathbf{\beta}$ matrix contains the path coefficients making up the Between indirect effect:

$$\mathbf{\beta} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{MX} & 0 & 0 \\ \beta_{YX} & \beta_{YM} & 0 \end{bmatrix} \tag{B4}$$

The total indirect effect of X_j on Y_j via M_{ij} is given by

$$\mathbf{F} = (\mathbf{I} - \mathbf{\beta})^{-1} - \mathbf{I} - \mathbf{\beta} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{MX}\beta_{YM} & 0 & 0 \end{bmatrix}, \tag{B5}$$

in which $\beta_{MX}\beta_{YM}$ is the Between indirect effect of the first column variable (X_j) on the last row variable (the Between portion of Y_{ij}) via the cluster-level M_j.

(Appendices continue)

Mediation in 1-1-1 Designs

If all three variables involved in the mediation model are assessed at Level 1, the design is termed 1-1-1. Krull and MacKinnon (2001) and Pituch et al. (2005) investigated forms of this model with random intercepts but fixed slopes. To fit the random-slopes model for 1-1-1 data as a special case of the general framework, the model reduces to the equations

$$
\mathbf{Y}_{ij} = \begin{bmatrix} X_{ij} \\ M_{ij} \\ Y_{ij} \end{bmatrix} = \mathbf{\Lambda}\mathbf{\eta}_{ij} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix}
\tag{B6}
$$

$$
\mathbf{\eta}_{ij} = \begin{bmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ B_{MXj} & 0 & 0 & 0 & 0 & 0 \\ B_{YXj} & B_{YMj} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Xij} \\ \zeta_{Mij} \\ \zeta_{Yij} \\ 0 \\ 0 \\ 0 \end{bmatrix}
\tag{B7}
$$

$$
\mathbf{\eta}_{j} = \begin{bmatrix} B_{MXj} \\ B_{YXj} \\ B_{YMj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{BMXj} \\ \mu_{BYXj} \\ \mu_{BYMj} \\ \mu_{\alpha\eta Xj} \\ \mu_{\alpha\eta Mj} \\ \mu_{\alpha\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & B_{MX} & 0 & 0 \\ 0 & 0 & 0 & B_{YX} & B_{YM} & 0 \end{bmatrix} \begin{bmatrix} B_{MXj} \\ B_{YXj} \\ B_{YMj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{BMXj} \\ \zeta_{BYXj} \\ \zeta_{BYMj} \\ \zeta_{\alpha\eta Xj} \\ \zeta_{\alpha\eta Mj} \\ \zeta_{\alpha\eta Yj} \end{bmatrix}.
\tag{B8}
$$

Here, because X_{ij}, M_{ij}, and Y_{ij} all have Within and Between variation, both Within and Between indirect effects can be estimated. As in the model for 2-1-1 data, the 3×3 $\mathbf{\beta}_B$ submatrix of $\mathbf{\beta}$ contains the coefficients making up the Between indirect effect, given by

$$
\mathbf{F} = (\mathbf{I} - \mathbf{\beta}_B)^{-1} - \mathbf{I} - \mathbf{\beta}_B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{MX}\beta_{YM} & 0 & 0 \end{bmatrix},
\tag{B9}
$$

in which $\beta_{MX}\beta_{YM}$ is the Between indirect effect of the first column variable (the Between portion of X_{ij}) on the third row variable (the Between portion of Y_{ij}) via the cluster-level component of M_{ij}. Additionally, the matrix **B** contains the coefficients making up the Within indirect effect, and these elements may be either fixed slopes or the means of random slopes. Specifying the Within slopes B_{MXj} and B_{YMj} as fixed or random will determine how to quantify the Within indirect effect. If both μ_{BMXj} and μ_{BYMj} are the means of random slopes, the indirect effect is $\mu_{BMXj}\mu_{BYMj} + \psi_{\zeta BMXj,\zeta BYMj}$, where $\psi_{\zeta BMXj,\zeta BYMj}$ is the cluster-level covariance of the random slopes B_{MXj} and B_{YMj}. Thus, counterintuitively, even if either μ_{BMXj} or μ_{BYMj} (or both) is zero, the indirect effect may still be nonzero if $\psi_{\zeta BMXj,\zeta BYMj} \neq 0$. It is also conceivable for $\mu_{BMXj}\mu_{BYMj}$ and $\mu_{BMXj}\mu_{BYMj} + \psi_{\zeta BMXj,\zeta BYMj}$ to be of opposite signs if $\mu_{BMXj}\mu_{BYMj}$ and $\psi_{\zeta BMXj,\zeta BYMj}$ are of opposite signs and $|\psi_{\zeta BMXj,\zeta BYMj}| > |\mu_{BMXj}\mu_{BYMj}|$. If either B_{MXj} or B_{YMj} is fixed, the Within indirect effect is $\mu_{BMXj}\mu_{BYMj}$. The multilevel mediation model of Bauer et al. (2006) may be expressed as a special case of the general model described here by constraining $\mathbf{B} = \mathbf{\beta}$.

(Appendices continue)

Mediation in 1-1-2 Designs

If X_{ij} and M_{ij} are assessed at Level 1 but Y_j is assessed at Level 2, the design is termed 1-1-2. If we wish to fit the model for 1-1-2 data, the general MSEM may be constrained to yield the equations

$$\mathbf{Y}_{ij} = \begin{bmatrix} X_{ij} \\ M_{ij} \\ Y_j \end{bmatrix} = \mathbf{\Lambda}\boldsymbol{\eta}_{ij} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} \tag{B10}$$

$$\boldsymbol{\eta}_{ij} = \begin{bmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ B_{MXj} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Xij} \\ \zeta_{Mij} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{B11}$$

$$\boldsymbol{\eta}_j = \begin{bmatrix} B_{MXj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{BMXj} \\ \mu_{\alpha \eta Xj} \\ \mu_{\alpha \eta Mj} \\ \mu_{\alpha \eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta_{MX} & 0 & 0 \\ 0 & \beta_{YX} & \beta_{YM} & 0 \end{bmatrix} \begin{bmatrix} B_{MXj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{BMXj} \\ \zeta_{\alpha \eta Xj} \\ \zeta_{\alpha \eta Mj} \\ \zeta_{\alpha \eta Yj} \end{bmatrix}. \tag{B12}$$

The matrix $\boldsymbol{\beta}$ contains the slopes making up the Between indirect effect, given by $\beta_{MX}\beta_{YM}$, which is the Between indirect effect of the first column variable (the Between portion of X_{ij}) on the third row variable (the Between portion of Y_j) via the cluster-level component of M_{ij}.

Mediation in 1-2-2 Designs

If only X_{ij} is assessed at Level-1, the design is 1-2-2. A hypothetical research question for 1-2-2 data is whether firm-level safety climate mediates the relationship between individual employees' safety attitudes and firm-level injury rate. To fit this model in the general MSEM framework, the general model reduces to

$$\mathbf{Y}_{ij} = \begin{bmatrix} X_{ij} \\ M_j \\ Y_j \end{bmatrix} = \mathbf{\Lambda}\boldsymbol{\eta}_{ij} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Xij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} \tag{B13}$$

$$\boldsymbol{\eta}_{ij} = \begin{bmatrix} \eta_{Xij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Xij} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{B14}$$

$$\boldsymbol{\eta}_j = \begin{bmatrix} \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{\alpha \eta Xj} \\ \mu_{\alpha \eta Mj} \\ \mu_{\alpha \eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ \beta_{MX} & 0 & 0 \\ \beta_{YX} & \beta_{YM} & 0 \end{bmatrix} \begin{bmatrix} \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha \eta Xj} \\ \zeta_{\alpha \eta Mj} \\ \zeta_{\alpha \eta Yj} \end{bmatrix}. \tag{B15}$$

Because both of the dependent variables are assessed at Level 2, there can be no random slopes. Thus, the indirect effect of the Between portion of X_j on Y_{ij} via M_{ij} is $\beta_{MX}\beta_{YM}$.

Mediation in 2-1-2 Designs

If X_j and Y_j are assessed at Level 2 but M_{ij} is assessed at Level 1, the design is termed 2-1-2. For instance, Kozlowski and Klein (2000) described an example in which organization-level human resources practices are causally antecedent to organizational performance through their effect on individual citizenship behaviors. When we fit the model for 2-1-2 data as a special case of the general MSEM framework with a random intercept for M_{ij}, the general MSEM reduces to the equations

$$\mathbf{Y}_{ij} = \begin{bmatrix} X_j \\ M_{ij} \\ Y_j \end{bmatrix} = \Lambda\boldsymbol{\eta}_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Mij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} \tag{B16}$$

$$\boldsymbol{\eta}_{ij} = \begin{bmatrix} \eta_{Mij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Mij} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{B17}$$

$$\boldsymbol{\eta}_j = \begin{bmatrix} \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{\alpha\eta Xj} \\ \mu_{\alpha\eta Mj} \\ \mu_{\alpha\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ \beta_{MX} & 0 & 0 \\ \beta_{YX} & \beta_{YM} & 0 \end{bmatrix} \begin{bmatrix} \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha\eta Xj} \\ \zeta_{\alpha\eta Mj} \\ \zeta_{\alpha\eta Yj} \end{bmatrix}. \tag{B18}$$

The matrix $\boldsymbol{\beta}$ contains the coefficients making up the Between indirect effect, given by $\beta_{MX}\beta_{YM}$.

Mediation in 1-2-1 Designs

If X_{ij} and Y_{ij} are assessed at Level 1 but M_j is assessed at Level 2, the design is termed 1-2-1 (see, e.g., McDonald, 1994). In the general MSEM framework described here, the model (with a random slope linking X_{ij} and Y_{ij}) reduces to

$$\mathbf{Y}_{ij} = \begin{bmatrix} X_{ij} \\ M_j \\ Y_{ij} \end{bmatrix} = \Lambda\boldsymbol{\eta}_{ij} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Xij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} \tag{B19}$$

$$\boldsymbol{\eta}_{ij} = \begin{bmatrix} \eta_{Xij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ B_{YXj} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{Xij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Xij} \\ \zeta_{Yij} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{B20}$$

$$\boldsymbol{\eta}_j = \begin{bmatrix} B_{YXj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{BYXj} \\ \mu_{\alpha\eta Xj} \\ \mu_{\alpha\eta Mj} \\ \mu_{\alpha\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta_{MX} & 0 & 0 \\ 0 & \beta_{YX} & \beta_{YM} & 0 \end{bmatrix} \begin{bmatrix} B_{YXj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{BYXj} \\ \zeta_{\alpha\eta Xj} \\ \zeta_{\alpha\eta Mj} \\ \zeta_{\alpha\eta Yj} \end{bmatrix}. \tag{B21}$$

The matrix $\boldsymbol{\beta}$ contains the path coefficients making up the Between indirect effect of the cluster-level component of X_{ij} on the cluster-level component of Y_{ij} via M_j, given by $\beta_{MX}\beta_{YM}$.