Structural Equation Modeling Approaches for Analyzing Partially Nested Data

Sonya K. Sterba and Kristopher J. Preacher

Department of Psychology and Human Development, Vanderbilt University

Rex Forehand

Department of Psychological Science, University of Vermont

Emily J. Hardcastle, David A. Cole, and Bruce E. Compas

Department of Psychology and Human Development, Vanderbilt University

Study designs involving clustering in some study arms, but not all study arms, are common in clinical treatment-outcome and educational settings. For instance, in a treatment arm, persons may be nested in therapy groups, whereas in a control arm there are no groups. Methodological approaches for handling such partially nested designs have recently been developed in a multilevel modeling framework (MLM-PN) and have proved very useful. We introduce two alternative structural equation modeling (SEM) approaches for analyzing partially nested data: a multivariate single-level SEM (SSEM-PN) and a multiple-arm multilevel SEM (MSEM-PN). We show how SSEM-PN and MSEM-PN can produce results equivalent to existing MLM-PNs and can be extended to flexibly accommodate several modeling features that are difficult or impossible to handle in MLM-PNs. For instance, using an SSEM-PN or MSEM-PN, it is possible to specify complex structural models involving cluster-level outcomes, obtain absolute model fit, decompose person-level predictor effects in the treatment arm using latent cluster means, and include traditional factors as predictors/outcomes. Importantly, implementation of such features for partially nested designs differs from that for fully nested designs. An empirical example involving a partially nested depression intervention combines several of these features in an analysis of interest for treatment-outcome studies.

A common design in treatment-outcome studies involves clustering in one or more study arms but not in other arm(s); such designs can be called *partially nested*. Whereas in fully nested studies groups may be randomized to condition, in partially nested studies individuals may be randomly assigned to condition, and then groups (clusters) may be constructed in the treatment arm for the purpose of treatment administration. Many such partially nested studies compare group therapy treatment (with persons nested within therapy groups) versus a wait-list or medication-only control condition (e.g., Compas et al., 2009, 2011; Dannon, Gon-Usishkin, Gelbert, Lowengrub, & Grunhaus, 2004; Haugli, Steen,

Laerum, Nygard, & Finset, 2001; Heller-Boersma, Schmidt, & Edmonds, 2007; Jarrett et al., 1999; Ladouceur et al., 2000; Pisinger, Vestbø, Borch-Johnsen, & Jørgensen, 2005; Price & Anderson, 2011; Thomas et al., 2010; Thompson, Gallagher, & Breckenridge, 1987; Van Minnen, Hoogduin, Keijsers, Hellenbrand, & Hendriks, 2003). Other examples involve educational studies comparing problem-solving outcomes in individual versus collaborative team environments (Kirschner, Paas, Kirschner, & Janssen, 2011; Vadasy & Sanders, 2008) or public health studies comparing outcomes from patients assigned to individual versus paired HIV counseling (Becker, Mlay, & Schwandt, 2010). Indeed, a recent review found partially nested designs used in approximately one third of randomized trials (Bauer, Sterba, & Hallfors, 2008).

In such partially nested designs, the common assumption of independently distributed observations is violated

for observations in the clustered arm(s). Outcomes within a clustered arm may be correlated because, for instance in an educational intervention, fast-learning readers interact with and elevate the performance of other students in their cluster. Correlated outcomes in some study arms complicate analysis of treatment effects and necessitate the explanation of variability both within and across clusters. The analytical challenges in the partial nesting context differ from those in other designs with naturally occurring singletons (single-person clusters), such as when all but one cluster member are absent due to subject-initiated missing data. In the partial nesting context, the singletons' presence is systematically paired with a covariate (study arm), and different model-implied variances and/or effects are theoretically anticipated in the nested versus nonnested arms (Sanders, 2011).

Methodologists have begun to attend to the analysis of data from such partially nested designs. Bauer et al. (2008), Lee and Thompson (2005), Moerbeek and Wong (2008), and Roberts and Roberts (2005) adapted a multilevel model for partially nested data (MLM-PN)—accommodating nesting only in the clustered study arm(s) but not in the unclustered study arm(s). In simulation studies, others have compared this approach with alternatives involving (a) ignoring the clustering entirely or (b) pretending all observations are clustered with the same intraclass correlation (ICC) and applying a standard multilevel model (Baldwin, Bauer, Stice, & Rohde, 2011; Korendijk, Maas, Hox, & Moerbeek, 2012; Sanders, 2011). The MLM-PN's estimates were unbiased and it outperformed alternatives (a) and (b) in terms of efficiency and Type I error. Comparatively, standard errors of the treatment effect were biased for alternatives (a) and (b), and variance component estimates were biased and uninterpretable for alternative (b) (Korendijk et al., 2012).[1] Other recent work has used this MLM-PN approach in sample size planning for optimal power (Candel & Van Breukelen, 2009, 2010; Moerbeek & Wong, 2008). In sum, this MLM-PN approach has proved very useful to date.

Social science and health researchers conducting treatment-outcome studies, however, often are interested in capitalizing on the flexibility of a structural equation modeling (SEM) framework when evaluating complex hypotheses; these researchers may also be interested in evaluating absolute model fit (Beran & Violato, 2010; Cole, Maxwell, Arvey, & Salas, 1993; Hoyle & Smith, 1994; Russell, Kahn, Spoth, & Altmaier, 1998; Tomarken & Waller, 2005). Researchers conducting partially nested treatment-outcome studies are

no exception. In particular, extending the MLM-PN to accommodate complex structural relations among multivariate outcomes can be less straightforward; including cluster-level outcomes or traditional factor predictors/outcomes is not possible, and there are limited options for decomposing person-level predictor effects into between- and within-cluster effects.

To address this need, we introduce two flexible SEM approaches for handling partially nested data. We show that both are statistically equivalent to the existing MLM-PN—when the MLM-PN is applicable. One SEM approach, a multivariate single-level SEM (SSEM-PN), can be implemented in standard SEM software. The other SEM approach involves a multiple-arm[2] multilevel SEM (MSEM-PN). All approaches require raw data and accommodate subject-initiated missing data on outcomes in any study arm under the missing-at-random assumption of full information maximum likelihood (FIML). Particular patterns of design-induced missingness are a by-product of the partial nesting design itself (described later). When estimating SSEM-PNs or MSEM-PNs, SEM software may provide absolute fit indices by default; however, these are untrustworthy. They are computed using a likelihood for a default saturated model that does not include relevant design-based constraints. Computing interpretable absolute fit indices for nonsaturated SSEM-PNs and MSEM-PNs that do not employ random slopes requires separate specification of a saturated model that accounts for design-based constraints: exchangeability of persons within cluster in the treatment arm (following Bauer, 2003) and independence of persons across arms. Two examples are given later. For all approaches–MLM-PN, SSEM-PN, and MSEM-PN—model selection can also involve information criteria and likelihood ratio difference tests.

In the first part of this article, we review the MLM-PN. Then we introduce the SSEM-PN and MSEM-PN in equations and path diagrams. We explain how these SEM approaches can equivalently represent each feature of the MLM-PN approach, including residual variances that are homoscedastic versus heteroscedastic across study arm, individual-level and cluster-level predictors, and multiple study arms. Simulated examples illustrate the equivalency of the SEM alternatives with the MLM-PN. After establishing this equivalency, we turn to two illustrative contexts where the added flexibility of the SEM approaches poses benefits: (a) when including cluster-level outcomes among multivariate outcomes and (b) when interest is in decomposing person-level predictor effects into within- and between-cluster components in the treatment arm. Both of these contexts are later incorporated into an empirical example analysis using a partially nested depression intervention

---

[1]Some of these simulations also compared the MLM-PN with other alternatives including (a) treatment clusters as fixed effects; (b) assigning all control persons the same subject identifier and considering them as constituting a single large cluster when fitting a standard MLM; or (c) after study completion, dividing control persons arbitrarily into multiperson "clusters" to mimic the data structure in the treatment arm and then fitting a standard MLM. The MLM-PN also performed favorably vis-à-vis these alternatives (Baldwin et al. 2011; Korendijk et al., 2012; Sanders, 2011).

[2]We use the term "multiple-*arm* model" in place of the more conventional SEM term "multiple-group model" because we have already used the word "group" to refer to cluster (as in therapy group).

design. Discussion focuses on considerations when choosing among MLM-PN, SSEM-PN, and MSEM-PN in practice, as well as further modeling possibilities. For all fitted SSEM-PNs and MSEM-PNs, Mplus (L. K. Muthén & Muthén, 1998–2013) syntax is given in an online Appendix; LIS-REL (Jöreskog & Sörbom, 2006) syntax is also provided for select examples.[3] Although we use a conventional matrix representation for the SSEM-PN, we present the MLM-PN and MSEM-PN in scalar format for simplicity. Corresponding matrix representations are provided in our online Appendix at www.vanderbilt.edu/peabody/sterba/appxs.htm

REVIEW OF THE MULTILEVEL MODELING APPROACH FOR HANDLING PARTIAL NESTING (MLM-PN)

We begin by reviewing the multilevel modeling approach for handling partial nesting (MLM-PN; Bauer et al., 2008; Lee & Thompson, 2005; Moerbeek & Wong, 2008; Roberts & Roberts, 2005) in which the number of arms, A, equals 2. Here the treatment arm is clustered and the control arm is unclustered; however, we can alternatively think of each person in the control arm as constituting his or her own cluster. Key features of the MLM-PN are the allowance for different model-implied variances in the control and treatment arms and allowance for separation of between-cluster and within-cluster outcome variability only in the treatment arm. To accomplish this, a random effect for a treatment indicator (coded $treat = 0$ for controls, $treat = 1$ for treated) is used in lieu of a random intercept to account for clustering. When $treat = 1$, the model includes a single random effect allowing for correlated observations within cluster, but when $treat = 0$, this random effect drops out of the model:

$$\text{Level 1: } y_{ij} = \beta_{0j} + \beta_{1j}treat_{ij} + r_{ij} \qquad (1)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\text{Reduced form: } y_{ij} = \gamma_{00} + \gamma_{10}treat_{ij} + u_{1j}treat_{ij} + r_{ij}.$$

y_{ij} is the outcome for person i (where $i = 1...N_j$) in cluster j (where $j = 1...J$). J is the total number of clusters. Hence, J is the sum of all (multiple-person) clusters in the treatment arm plus all (singleton) clusters in the control arm. N_j is the number of persons in the jth cluster. Usually $N_j > 1$ in the treatment arm (though fewer could arise in practice due to subject-initiated missing data) but $N_j = 1$ in the control arm. The total number of persons in the sample is N. β_{0j} is the latent intercept for cluster j, with mean γ_{00} interpreted as the

mean in the control arm; it is fixed across clusters. β_{1j} is the effect of treatment (i.e., difference between control and treatment) for cluster j, with mean γ_{10} and cluster-specific deviation u_{1j}. β_{1j} differs across clusters only when $treat = 1$ (otherwise the $u_{1j}treat_{ij}$ term drops out of the reduced form expression). Cluster-specific deviations are assumed normally distributed:

$$u_{1j} \sim N(0, \tau_{11}). \qquad (2)$$

τ_{11} is the between-cluster variance in the treatment arm. r_{ij} is a residual, and r_{ij} and u_{1j} are assumed independent. r_{ij} are also assumed independent across study arms by design and may have either homoscedastic (Equation (3a)) or heteroscedastic (Equation (3b)) variance across arms:

$$r_{ij} \sim N(0, \sigma^2) \qquad (3a)$$

$$r_{ij}|treat_{ij} = 1 \sim N(0, \sigma_t^2) \qquad (3b)$$

$$r_{ij}|treat_{ij} = 0 \sim N(0, \sigma_c^2).$$

Imposing homoscedastic residual variances forces the model-implied variance in the treatment arm ($\tau_{11} + \sigma^2$) to be greater than or equal to that in the control arm (σ^2), which may not be plausible if the cluster-based treatment administration homogenizes outcomes via contagion (Roberts & Roberts, 2005). Heteroscedastic residual variances across arms allows for this possibility, permitting $\sigma_c^2 > \sigma_t^2$.

Substantively, note that for persons in the control arm σ^2 (or σ_c^2 under heteroscedasticity) does not represent "within-cluster variance" per se, as there is no splitting of variance into within- and between-cluster components, and each person can be conceptualized as constituting a cluster. Rather, this parameter simply represents the variance in the control arm, whereas in the treatment arm σ^2 (or σ_t^2 under heteroscedasticity) does represent within-cluster variance.

Adding Exogenous Covariates

Person-level (Level 1) covariates such as patient's duration of illness (e.g., x_{ij}) have also been included in the MLM-PN. As noted by Bauer et al. (2008), x_{ij} could potentially vary at both levels in the treatment arm because it can be decomposed into a cluster j average and an individual deviation from the cluster j average. Nonetheless, previous examples using MLM-PN with person-level covariates (e.g., Baldwin et al., 2011; Bauer et al., 2008; Moerbeek & Wong, 2008; Roberts & Roberts, 2005) have estimated a single (total) effect for x_{ij}, as in Equation (4).

The Equation (4) MLM-PN includes both a Level 1 x_{ij} and a Level 2 w_j (the latter is discussed shortly).

$$\text{Level 1: } y_{ij} = \beta_{0j} + \beta_{1j}treat_{ij} + \beta_{2j}x_{ij} + r_{ij} \qquad (4)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_j + u_{1j}$$

$$\beta_{2j} = \gamma_{20}.$$

---

[3]Equations in the text correspond with a conditional likelihood formulation (e.g., as used in Mplus). Implementation with LISREL (which uses a joint likelihood formulation) requires slightly different specification for some models; examples are given in the online Appendix.

Reduced form:

$$y_{ij} = \gamma_{00} + \gamma_{10}treat_{ij} + \gamma_{11}treat_{ij}w_j$$
$$+ \gamma_{20}x_{ij} + u_{1j}treat_{ij} + r_{ij}.$$

γ_{20} is a total effect of x_{ij} in the treatment arm but a simple effect in the control arm. Other options for incorporating x_{ij} are deferred to a separate section, after presenting MLM-PN, SSEM-PN, and MLM-PN. For now we employ the previously used practice (total effect of x_{ij}). Note that allowing interactions of predictors (e.g., x_{ij}) with treatment in MLM-PN requires construction and addition of product terms. SSEM-PN and MSEM-PN will not require this step.

Including a cluster-level (Level 2) covariate w_j in MLM-PN can present some complexities that were addressed by Bauer et al. (2008) and Lee and Thompson (2005). The complexities arise when cluster-level predictors (e.g., group cohesion) are undefined, and thus missing-by-design, for the control arm. However, a w_j measured at the cluster level in the treatment arm may sometimes *also* be measured in the control arm—where it is synonymously a cluster-level or person-level predictor. For instance, if a treatment arm received classroom-based Scholastic Aptitude Test (SAT) instruction and a comparison arm received individual SAT tutoring, teacher experience could be considered a cluster-level predictor measured in both arms. Unless otherwise stated, here we focus on the situation in which cluster-level predictors are missing-by-design in the control arm because this situation presents unique complexities. In this situation, MLM-PN includes an effect, γ_{11}, for the interaction of $treat_{ij}$ and w_j (but not a main effect of w_j) so that the $treat_{ij}w_j$ term drops out of the reduced-form expression for controls, as shown in Equation (4). Also, to prevent all control participants from being listwise deleted due to missingness-by-design on w_j (a typical consequence of model fitting with FIML using a likelihood conditional on predictors), Bauer et al. (2008) assigned them an arbitrary value on w_j. This assigned value is not a missing data code; we recommend using 0 because a large number can cause estimation instability. This assignment does not affect results because the interaction term drops out of the reduced form expression for controls.

Adding Study Arms

Finally, adding a third study arm ($A = 3$) corresponding to a second treatment condition is accomplished in MLM-PN by forming two dummy treatment indicator variables $treat1_{ij}$ and $treat2_{ij}$ rather than one and entering both as person-level predictors. If there is nesting in the third study arm (e.g., a couples-therapy arm in addition to the original group-therapy and control arms), a random effect, u_{3j}, would be included for $treat2_{ij}$ also, as in

Equation (5).

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}treat1_{ij} + \beta_{2j}x_{ij} + \beta_{3j}treat2_{ij}$
$$+ r_{ij} \qquad (5)$$

Level 2: $\beta_{0j} = \gamma_{00}$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_j + u_{1j}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31}w_j + u_{3j}.$$

Reduced form:

$$y_{ij} = \gamma_{00} + (\gamma_{10} + \gamma_{11}w_j + u_{1j})treat1_{ij} + \gamma_{20}x_{ij}$$
$$+ (\gamma_{30} + \gamma_{31}w_j + u_{3j})treat2_{ij} + r_{ij}.$$

Again, residual variances can be homoscedastic[4] (σ^2), or heteroscedastic (σ_c^2 vs. $\sigma_{t_1}^2$ vs. $\sigma_{t_2}^2$), across arms. Because of the nature of the design, random effects for separate arms are assumed independent:

$$\begin{bmatrix} u_{1j} \\ u_{3j} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{11} & \\ 0 & \tau_{33} \end{bmatrix} \right). \qquad (6)$$

If the effect of w_j is conjectured to be the same in the couples- and group-therapy treatment arms, an equality constraint $\gamma_{11} = \gamma_{31}$ could be imposed. If there is no nesting in the third study arm (e.g., if an individual-therapy arm were added), then u_{3j} and τ_{33} are omitted. If w_j (e.g., therapist experience) is a Level 2 predictor in the clustered, group-therapy arm but is also measured in the individual therapy arm, an interaction of $treat2_{ij}$ and w_j can still be included in the reduced form expression.

MULTIVARIATE SINGLE-LEVEL SEM APPROACH FOR HANDLING PARTIAL NESTING (SSEM-PN)

In the treatment arm of the MLM-PN, a single outcome was measured cross-sectionally on multiple exchangeable persons (Level 1 units) nested within person-group clusters (e.g., therapy groups). In contrast, cross-sectional applications of SEM measurement models typically involve multiple different outcomes (Level 1 units) nested in a person (Level 2 unit), where the multiple different outcomes serve as indicators of common factor(s). However, an SEM measurement model can instead be applied to a *single* outcome measured cross-sectionally on *multiple* exchangeable persons (Level 1 units) nested within person-group clusters (Level 2 units); then the outcomes for all persons in a cluster load equally on a common factor (see, e.g., Bauer, 2003; Mehta & Neale, 2005).

---

[4]In practice, homoscedasticity of residual variances across arm is more commonly considered for unconditional models or when the same Level 1 predictors are used in each arm.

The multivariate single-level SEM for partial nesting (SSEM-PN) uses the latter approach in each study arm. In each arm, the outcome of each person in a cluster loads on a common intercept factor (i.e., random intercept) with unit loadings and equal residual variances within cluster. In the treatment arm, this intercept factor has multiple indicators (one per person), but in the control arm the intercept factor has a single indicator. Because the intercept factor has multiple indicators in the treatment arm, the outcome variance can be partitioned into between- versus within-cluster components. Because the intercept factor has only one indicator in the control arm, the outcome variance cannot be partitioned. Importantly, this SSEM-PN cannot be specified using a conventional multiple-arm SEM (Jöreskog, 1971) because this would require the same outcomes to be used in both arms and would not allow outcomes to be completely missing for one arm (e.g., Kim, Mun, & Smith, in press; Widaman, Grimm, Early, Robins, & Conger, 2013).[5] Control arm persons have scores missing-by-design on all but the first y-variable. Instead, here the SSEM-PN specification is facilitated by creating a modified wide-format data set that allows fitting a different model per arm using different numbers of variables per arm. We begin by contrasting the data structure between the MLM-PN and SSEM-PN before turning to the equations used to specify the SSEM-PN.

For the MLM-PN, data were structured in long format, as shown in Figure 1. In long format, we used N rows (such that each row is a person) and $1 + p + q$ columns: 1 is the total number of outcome variables, p is the number of person-level predictors, and q is the number of cluster-level predictors. SSEM-PN instead requires a modified wide-format data set, also in Figure 1, in which each of J rows corresponds to a cluster (of ≥ 1 persons). For the two-arm example, $1 + n_{t_1}$ columns represent the outcome variable, 1 column represents an exogenous cluster-level covariate, and n_{t_1} columns represent an exogenous person-level covariate. $n_{t_1} = \max_{t_1}(N_j)$, which is 5 here in our illustration.

Our basic SSEM-PN, where the only predictor is treatment, uses a subset of matrices in the general SEM model of, for example, B. O. Muthén (2002; see also B. O. Muthén,

---

[5]One approach to circumvent this limitation is described shortly in this article; it involves minor recoding of the outcome variable but in exchange affords the benefits of (a) a transparent model specification in which all parameters are interpretable and (b) a likelihood identical to equivalent MSEM-PNs and MLM-PNs fitted in this article. Some software packages (e.g., Mplus but not LISREL) permit an alternative approach relying on a true multiple-arm SEM architecture, but require a trick in which variables for nonexistent persons are nonetheless included in the control arm model specification, with their variances fixed to near 0 and their effects held equal to those in the treatment arm. This approach can encounter estimation problems for some of the more complex models considered later. It also entails a less transparent model specification (i.e., half of the specified model is not interpretable). For these reasons, the latter approach was not used in the SSEM-PN. Other approaches are discussed in Widaman et al. (2013) and Kim et al. (in press) in different contexts (e.g., measurement invariance testing).

1987). Matrices currently used in Equation (7a) have their elements defined in (7b):

$$\mathbf{y}_j = \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \boldsymbol{\varepsilon}_j$$
$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\zeta}_j \qquad (7a)$$

$$
\begin{bmatrix} y^c_{1j} \\ y^t_{1j} \\ y^t_{2j} \\ y^t_{3j} \\ y^t_{4j} \\ y^t_{5j} \end{bmatrix}
=
\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}
\begin{bmatrix} \eta^c_j \\ \eta^t_j \end{bmatrix}
+
\begin{bmatrix} \varepsilon^c_{1j} \\ \varepsilon^t_{1j} \\ \varepsilon^t_{2j} \\ \varepsilon^t_{3j} \\ \varepsilon^t_{4j} \\ \varepsilon^t_{5j} \end{bmatrix}
\qquad (7b)
$$

$$
\begin{bmatrix} \eta^c_j \\ \eta^t_j \end{bmatrix}
=
\begin{bmatrix} \mu^c \\ \mu^t \end{bmatrix}
+
\begin{bmatrix} 0 \\ \zeta^t_j \end{bmatrix}
$$

y^c_{1j} is the sole outcome for a person (i.e., singleton cluster) in the control arm. y^t_{1j}-y^t_{5j} are outcomes for the first through last (here, fifth) persons in a treatment-arm cluster. This SSEM-PN is diagrammed in Figure 2, Panel A, for a generic cluster in the treatment arm (consisting of y^t_{1j}-y^t_{5j}) and a generic cluster in the control arm (consisting of y^c_{1j}). Because persons in a treatment-arm cluster are considered exchangeable, their ordering does not matter. If the jth cluster corresponds to a control cluster, y^t_{1j}-y^t_{5j} are coded as missing-by-design. If the jth cluster corresponds to a treatment cluster, y^c_{1j} is coded missing-by-design. Hence, $\mathbf{y}_j$ is a $(1+n_{t_1})$ by 1 vector of observed outcomes for cluster j.

$\boldsymbol{\eta}_j$ is an A by 1 vector of latent factors. Recall that A is the number of arms. η^c_j is the intercept factor in the control arm and η^t_j is the intercept factor in the treatment arm. $\boldsymbol{\Lambda}$ is a $(1+n_{t_1})$ by A design matrix. $\boldsymbol{\mu}$ is an A by 1 vector of factor means. η^c_j has mean μ^c, corresponding to γ_{00} from MLM-PN. η^t_j has mean μ^t, so μ^t-μ^c yields the average treatment effect, that is, γ_{10} from MLM-PN.

$\boldsymbol{\zeta}_j$ is an A by 1 vector of cluster-specific deviations from factor means, distributed as $\boldsymbol{\zeta}_j \sim N(\mathbf{0}, \boldsymbol{\Psi})$:

$$
\begin{bmatrix} 0 \\ \zeta^t_j \end{bmatrix}
\sim N \left(
\begin{bmatrix} 0 \\ 0 \end{bmatrix},
\begin{bmatrix} 0 & 0 \\ 0 & \psi^t \end{bmatrix}
\right). \qquad (8)
$$

$\boldsymbol{\varepsilon}_j$ is a $(1+ n_{t_1})$ by 1 vector of residuals for cluster j, which are independent across arm due to the study design and are distributed as $\boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \boldsymbol{\Theta}_\varepsilon)$, where

$$
\begin{bmatrix} \varepsilon^c_{1j} \\ \varepsilon^t_{1j} \\ \varepsilon^t_{2j} \\ \varepsilon^t_{3j} \\ \varepsilon^t_{4j} \\ \varepsilon^t_{5j} \end{bmatrix}
\sim N \left(
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},
\begin{bmatrix}
\theta^c_\varepsilon & & & & & \\
0 & \theta^t_\varepsilon & & & & \\
0 & 0 & \theta^t_\varepsilon & & & \\
0 & 0 & 0 & \theta^t_\varepsilon & & \\
0 & 0 & 0 & 0 & \theta^t_\varepsilon & \\
0 & 0 & 0 & 0 & 0 & \theta^t_\varepsilon
\end{bmatrix}
\right). \qquad (9)
$$

That is, in SSEM-PN's treatment arm, the cluster-specific deviation from the mean, ζ^t_j, has variance ψ^t, and the person-specific residual, ε^t_{ij}, has variance θ^t_ε. Hence, for the treatment

Long-format data structure
for MLM-PN or MSEM-PN

clusterID	y	tx
38	0.88	1
38	2.86	1
38	1.53	1
38	1.85	1
38	2.37	1
39	1.20	1
39	1.80	1
39	1.33	1
39	1.59	1
39	0.95	1
40	1.32	1
40	-2.04	1
40	-0.71	1
40	-1.15	1
40	-0.94	1
41	5.17	0
42	6.56	0
43	6.54	0
44	5.03	0

Modified wide-format data structure for SSEM-PN

clusterID	tx	y1c	y1t	y2t	y3t	y4t	y5t
38	1	.	0.88	2.86	1.53	1.85	2.37
39	1	.	1.20	1.80	1.33	1.59	0.95
40	1	.	1.32	-2.04	-0.71	-1.15	-0.94
41	0	5.17	.	.	.	.	.
42	0	6.56	.	.	.	.	.
43	0	6.54	.	.	.	.	.
44	0	5.03	.	.	.	.	.

FIGURE 1   Illustrative data set excerpt for a basic two-arm partial nesting model: long format (for MLM-PN/MSEM-PN) versus modified wide format (for SSEM-PN). *Note.* Data shown for three treatment-arm clusters (each containing 5 individuals) and for 4 individuals in the control arm (each labeled with his or her own clusterID: #41–44). "." = missing code; tx = treatment; SSEM-PN = multivariate single-level structural equation model (SEM) for partial nesting; MLM-PN = multilevel model for partial nesting; MSEM-PN = multiple arm multilevel SEM for partial nesting.

arm, outcome variance is partitioned into between-cluster, ψ^t, and within-cluster, θ_ε^t, components. ψ^t is equal to τ_{11} from MLM-PN and θ_ε^t matches σ_t^2 from MLM-PN. In SSEM-PN's control arm, we have the option of *either* estimating a variance for a person-specific residual ε_{1j}^c (i.e., θ_ε^c) *or* estimating a variance for a cluster-specific residual ζ_j^c (i.e., ψ^c)—but not both—in order to represent the outcome variance in the control arm. Equations (7)–(9) use the former option (as also illustrated in Figure 2, Panel A). We omit ζ_j^c, implying $\psi^c = 0$, and include ε_{1j}^c with variance θ_ε^c. Note θ_ε^c corresponds to σ_c^2 from MLM-PN. However, for completeness, we also depict the latter option in Figure 2, Panel B, because we use it later in more complex models. Both options (whether estimating ψ^c or θ_ε^c) would give the same estimate for the variance in the control arm. Finally, if homoscedastic residual variances are desired across treatment and control arms, we constrain $\theta_\varepsilon^c = \theta_\varepsilon^t$ in Equation (9), which then equals σ^2 from the MLM-PN.[6] Equation (9) depicts heteroscedastic residual variances across arm.

This basic two-arm homoscedastic residual variance SSEM-PN is not saturated and contains no random slopes. A researcher could examine its overall fit using absolute model fit indices (χ^2, RMSEA, etc.), computed using the likelihood from the fitted model and the likelihood from a mod-

ified saturated model[7] that incorporates design-based constraints: exchangeability of persons in the treatment arm (see Bauer, 2003) and independence of persons across arms. Such a modified saturated model could entail estimating means, variances, and covariances of y_{1j}^t-y_{5j}^t (each constrained to equality for y_{1j}^t-y_{5j}^t); estimating the mean and variance of y_{1j}^c; and not allowing y_{1j}^t-y_{5j}^t to covary with y_{1j}^c. This modified saturated model totals five free parameters, implying $\Delta df = 1$, and is equivalent to a two-arm heteroscedastic residual variance SSEM-PN. As we later mention in the Discussion, absolute fit for SSEM-PNs is of increasing interest when more theory-based constraints are imposed in the fitted model beyond those required by the design and for identification.

Adding Exogenous Covariates

Extending the SSEM-PN to include an exogenous person-level covariate and cluster-level covariate gives Equation (10a). Corresponding elements are given in Equation (10b).

$$\mathbf{y}_j = \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \mathbf{K}\mathbf{x}_j + \boldsymbol{\varepsilon}_j$$
$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\Omega}\mathbf{w}_j + \boldsymbol{\zeta}_j \qquad (10a)$$

---

[6]In the control arm, if the Level 2 variance (ψ^c) had instead been estimated, as in Figure 2, Panel B, achieving homoscedastic residual variances across arm would require the constraint: $\theta_\varepsilon^t = \psi^c$.

[7]Considerable subject-initiated missing data in a given arm, leading to low covariance coverage, can preclude calculation of the modified saturated model in Mplus (which does not happen when covariance coverage is exactly 0).
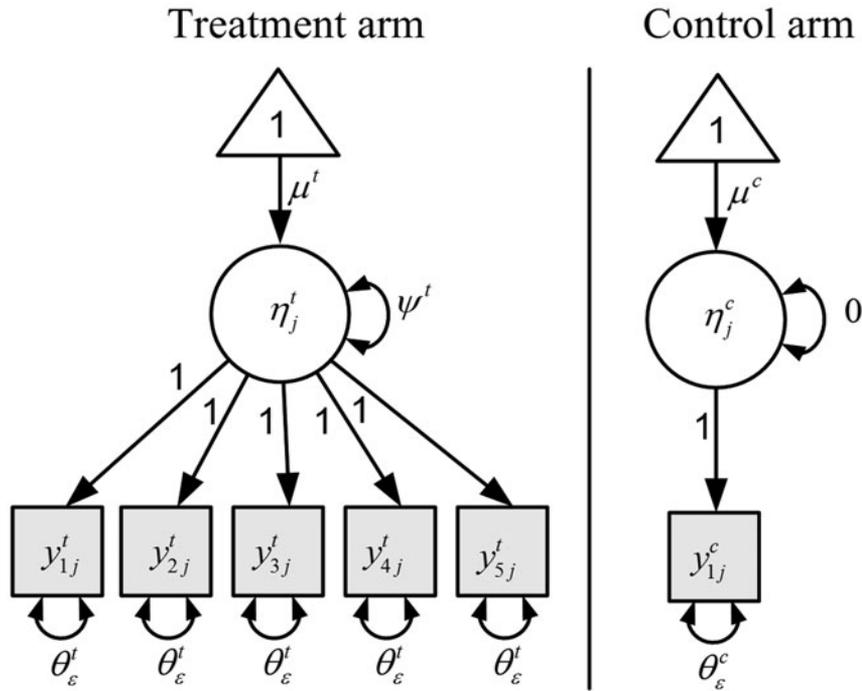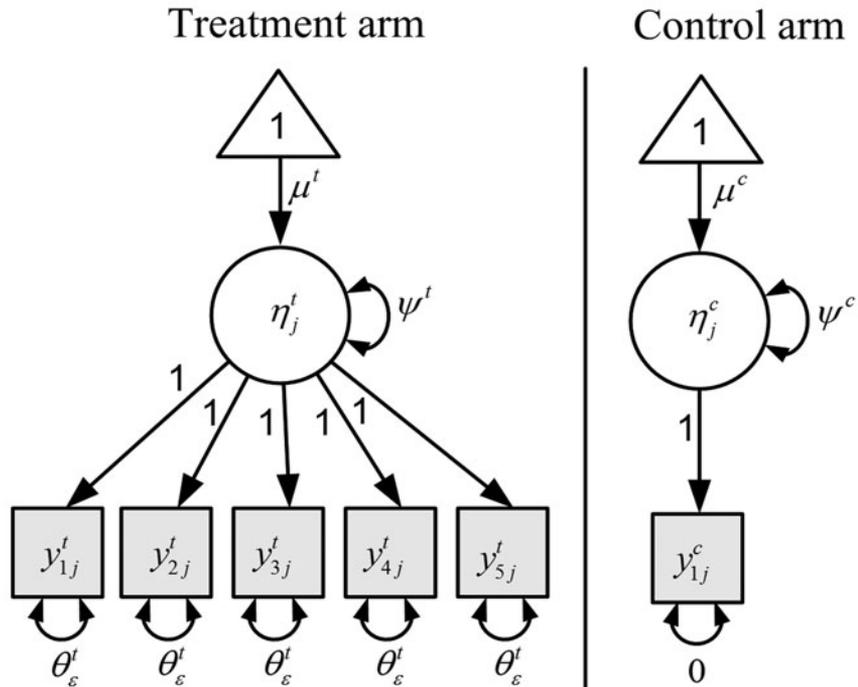
FIGURE 2   Multivariate single-level structural equation model for partial nesting (SSEM-PN). *Note.* In each panel, the diagram is drawn for the *j*th cluster in the treatment arm (left side) or control arm (right side) where in the control arm a person is his or her own cluster. Notation in Panel A corresponds to the two-arm SSEM-PN with heteroscedastic residual variances in Equations (7)–(9). To simplify presentation, residuals (ε_{ij}^t, ε_{ij}^c, ζ_j^t, ζ_j^c) are not shown; only their variances are shown. Measured variables are depicted by shaded boxes; latent variables by circles; constants by triangles; directed paths by straight arrows; and (co-)variances by curved arrows.

$$
\begin{bmatrix} y_{1j}^{c} \\ y_{1j}^{t} \\ y_{2j}^{t} \\ y_{3j}^{t} \\ y_{4j}^{t} \\ y_{5j}^{t} \end{bmatrix}
=
\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}
\begin{bmatrix} \eta_{j}^{c} \\ \eta_{j}^{t} \end{bmatrix}
$$

$$
+
\begin{bmatrix}
\kappa^{c} & 0 & 0 & 0 & 0 \\
\kappa^{t} & 0 & 0 & 0 & 0 \\
0 & \kappa^{t} & 0 & 0 & 0 \\
0 & 0 & \kappa^{t} & 0 & 0 \\
0 & 0 & 0 & \kappa^{t} & 0 \\
0 & 0 & 0 & 0 & \kappa^{t}
\end{bmatrix}
\begin{bmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ x_{4j} \\ x_{5j} \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{1j}^{c} \\ \varepsilon_{1j}^{t} \\ \varepsilon_{2j}^{t} \\ \varepsilon_{3j}^{t} \\ \varepsilon_{4j}^{t} \\ \varepsilon_{5j}^{t} \end{bmatrix}
\quad (10b)
$$

$$
\begin{bmatrix} \eta_{j}^{c} \\ \eta_{j}^{t} \end{bmatrix}
=
\begin{bmatrix} \mu^{c} \\ \mu^{t} \end{bmatrix}
+
\begin{bmatrix} 0 \\ \omega_{1}^{t} \end{bmatrix}
[w_{j}]
+
\begin{bmatrix} 0 \\ \zeta_{j}^{t} \end{bmatrix}
$$

$\mathbf{x}_{j}$ is a (pn_{t_1}) by 1 vector of exogenous person-level covariates; each person-level covariate is represented by n_{t_1} rows, here $x_{1j} - x_{5j}$. $\mathbf{K}$ is a $(1+n_{t_1})$ by (pn_{t_1}) matrix of regression coefficients for these Level 1 predictors. $\mathbf{w}_{j}$ is a q by 1 vector of cluster-level covariates, here containing a single w_{j}. $\mathbf{\Omega}$ is an A by q matrix of regression coefficients for cluster-level covariates. For control participants, $x_{2j} - x_{5j}$ and w_{j} are here missing-by-design.[8] ω_{1}^{t} corresponds to γ_{11} from the MLM-PN. Imposing the constraint $\kappa^{t} = \kappa^{c}$, yielding κ, corresponds to γ_{20} from the MLM-PN. This equality constraint was not imposed in Equation (10a) to highlight that an interaction of x_{ij} with treatment is allowed by simply relaxing this constraint.

Adding Study Arms

Adding a third study arm for a second treatment condition extends the rows of $\mathbf{y}_{j}$, $\mathbf{\Lambda}$, $\boldsymbol{\varepsilon}_{j}$, and $\mathbf{K}$ by n_{t_2} where $n_{t_2} = \max_{t_2}(N_j)$. Adding a third study arm also adds a column to $\mathbf{\Lambda}$ and a new row to $\boldsymbol{\eta}_{j}$, $\boldsymbol{\mu}$, $\mathbf{\Omega}$, and $\boldsymbol{\zeta}_{j}$. (Columns of $\mathbf{K}$ and rows of $\mathbf{x}_{j}$ also expand if $n_{t_2} > n_{t_1}$, which is not the case here.) For example, if the third study arm is nonnested (e.g.,

an individual therapy arm) then $n_{t_2} = 1$, and we have

$$
\begin{bmatrix} y_{1j}^{c} \\ y_{1j}^{t_1} \\ y_{2j}^{t_1} \\ y_{3j}^{t_1} \\ y_{4j}^{t_1} \\ y_{5j}^{t_1} \\ y_{1j}^{t_2} \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} \eta_{j}^{c} \\ \eta_{j}^{t_1} \\ \eta_{j}^{t_2} \end{bmatrix}
$$

$$
+
\begin{bmatrix}
\kappa^{c} & 0 & 0 & 0 & 0 \\
\kappa^{t_1} & 0 & 0 & 0 & 0 \\
0 & \kappa^{t_1} & 0 & 0 & 0 \\
0 & 0 & \kappa^{t_1} & 0 & 0 \\
0 & 0 & 0 & \kappa^{t_1} & 0 \\
0 & 0 & 0 & 0 & \kappa^{t_1} \\
\kappa^{t_2} & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ x_{4j} \\ x_{5j} \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{1j}^{c} \\ \varepsilon_{1j}^{t_1} \\ \varepsilon_{2j}^{t_1} \\ \varepsilon_{3j}^{t_1} \\ \varepsilon_{4j}^{t_1} \\ \varepsilon_{5j}^{t_1} \\ \varepsilon_{1j}^{t_2} \end{bmatrix}
\quad (11)
$$

$$
\begin{bmatrix} \eta_{j}^{c} \\ \eta_{j}^{t_1} \\ \eta_{j}^{t_2} \end{bmatrix}
=
\begin{bmatrix} \mu^{c} \\ \mu^{t_1} \\ \mu^{t_2} \end{bmatrix}
+
\begin{bmatrix} 0 \\ \omega_{1}^{t_1} \\ \omega_{1}^{t_2} \end{bmatrix}
[w_{j}]
+
\begin{bmatrix} 0 \\ \zeta_{j}^{t_1} \\ 0 \end{bmatrix}
$$

But if the third study arm is clustered, as in a couples therapy arm, then $n_{t_2} = 2$, as in Equation (12):

$$
\begin{bmatrix} y_{1j}^{c} \\ y_{1j}^{t_1} \\ y_{2j}^{t_1} \\ y_{3j}^{t_1} \\ y_{4j}^{t_1} \\ y_{5j}^{t_1} \\ y_{1j}^{t_2} \\ y_{2j}^{t_2} \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} \eta_{j}^{c} \\ \eta_{j}^{t_1} \\ \eta_{j}^{t_2} \end{bmatrix}
$$

$$
+
\begin{bmatrix}
\kappa^{c} & 0 & 0 & 0 & 0 \\
\kappa^{t_1} & 0 & 0 & 0 & 0 \\
0 & \kappa^{t_1} & 0 & 0 & 0 \\
0 & 0 & \kappa^{t_1} & 0 & 0 \\
0 & 0 & 0 & \kappa^{t_1} & 0 \\
0 & 0 & 0 & 0 & \kappa^{t_1} \\
\kappa^{t_2} & 0 & 0 & 0 & 0 \\
0 & \kappa^{t_2} & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ x_{4j} \\ x_{5j} \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{1j}^{c} \\ \varepsilon_{1j}^{t_1} \\ \varepsilon_{2j}^{t_1} \\ \varepsilon_{3j}^{t_1} \\ \varepsilon_{4j}^{t_1} \\ \varepsilon_{5j}^{t_1} \\ \varepsilon_{1j}^{t_2} \\ \varepsilon_{2j}^{t_2} \end{bmatrix}
\quad (12)
$$

$$
\begin{bmatrix} \eta_{j}^{c} \\ \eta_{j}^{t_1} \\ \eta_{j}^{t_2} \end{bmatrix}
=
\begin{bmatrix} \mu^{c} \\ \mu^{t_1} \\ \mu^{t_2} \end{bmatrix}
+
\begin{bmatrix} 0 \\ \omega_{1}^{t_1} \\ \omega_{1}^{t_2} \end{bmatrix}
[w_{j}]
+
\begin{bmatrix} 0 \\ \zeta_{j}^{t_1} \\ \zeta_{j}^{t_2} \end{bmatrix}.
$$

The conditional treatment effect in the third arm is μ^{t_2}-μ^{c} (the same as γ_{30} in MLM-PN). The effect of the person-level covariate in the third arm is κ^{t_2}. If the equality constraint $\kappa^{c} = \kappa^{t_1} = \kappa^{t_2}$ is imposed, the resultant κ is the same as γ_{20} from MLM-PN. Again, a treatment by covariate interaction is easily allowed by relaxing this constraint. The cluster-level covariate w_{j} may be undefined (and missing-by-design) for a nonnested third study arm; if so, $\omega_{1}^{t_2} = 0$ in Equation (11).

---

[8]A data management step similar to that used in MLM-PN (Bauer et al., 2008) is required to prevent listwise deletion of cases with covariate missingness-by-design (i.e., covariate missingness arising due to the partial nesting design structure) under FIML when using exogenous predictors and a conditional likelihood. Some SEM software uses a conditional likelihood (e.g., Mplus 6.1 or later but not LISREL 8.8); general documentation regarding this data management step is at www.statmodel.com. Here, for controls, an arbitrary value that is not the missing data code (we use 0) is assigned to $x_{2j} - x_{5j}$ and w_{j}. This insertion for $x_{2j} - x_{5j}$ does not affect results because corresponding outcomes $y_{2j}^{t} - y_{5j}^{t}$ are missing; this insertion for w_{j} does not affect results because here its slope for controls is fixed to 0.

Otherwise, the effect of w_j in the third arm is $\omega_1^{t_2}$, which is γ_{31} in the MLM-PN. A third arm expands dimensions of $\mathbf{\Psi}$ and $\mathbf{\Theta}_\varepsilon$:

$$\begin{bmatrix} 0 \\ \zeta_j^{t_1} \\ \zeta_j^{t_2} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 & \psi^{t_1} \\ 0 & 0 & \psi^{t_2} \end{bmatrix} \right) \quad (13)$$

$$\begin{bmatrix} \varepsilon_{1j}^c \\ \boldsymbol{\varepsilon}_j^{t_1} \\ \boldsymbol{\varepsilon}_j^{t_2} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \theta_\varepsilon^c \\ \mathbf{0} & \theta_\varepsilon^{t_1}\mathbf{I} \\ \mathbf{0} & \mathbf{0} & \theta_\varepsilon^{t_2}\mathbf{I} \end{bmatrix} \right) \quad (14)$$

where $\boldsymbol{\varepsilon}_j^{t_1}$ is $n_{t_1} \times 1$ and $\boldsymbol{\varepsilon}_j^{t_2}$ is $n_{t_2} \times 1$. Residuals and deviations are again assumed independent across study arm by design. Specifying homoscedastic[4] residual variances across arm requires the constraint $\theta_\varepsilon^c = \theta_\varepsilon^{t_1} = \theta_\varepsilon^{t_2}$ (corresponding to σ^2 from MLM-PN), and relaxing this constraint allows heteroscedastic residual variances (corresponding to σ_c^2, $\sigma_{t_1}^2$, and $\sigma_{t_2}^2$, respectively, from MLM-PN). If the third arm is clustered, ψ^{t_2} is the same as τ_{33} in MLM-PN. For a nonnested third arm, $\zeta_j^{t_2} = 0$ and $\psi^{t_2} = 0$.

MULTIPLE-ARM MULTILEVEL SEM APPROACH FOR HANDLING PARTIAL NESTING (MSEM-PN)

An alternative to adapting a multivariate single-level SEM for partial nesting is to adapt a multilevel SEM (MSEM) for partial nesting. In general, MSEM (e.g., Liang & Bentler, 2004; McDonald & Goldstein, 1989; B. O. Muthén, 1990; B. O. Muthén & Asparouhov, 2009) allows the decomposition of observed variables into latent between-cluster and within-cluster components. A *multiple-arm* MSEM allows different modeling of latent between-cluster and within-cluster components in each arm, and this feature is used in the MSEM-PN for accommodating the partial nesting structure. Like MLM-PN, but unlike SSEM-PN, it uses a long-format data structure (see Figure 1). In long format, here for both arms the outcome is still the univariate y_{ij} regardless of the size of the jth cluster. A path diagram of a basic two-arm MSEM-PN is represented in Figure 3, Panel A, for a generic *person* in each arm (in contrast to the Figure 2 SSEM-PN representation for a generic *cluster* in each arm). Again, in the control arm, each person is his or her own cluster. The corresponding MSEM-PN equation is

$$\begin{array}{c|c} \text{Treatment arm:} & \text{Control arm:} \\ y_{ij} = \tilde{y}_j^t + \tilde{y}_{ij}^t & y_{ij} = \tilde{y}_j^c + \tilde{y}_{ij}^c \\ \text{Between:} \quad \tilde{y}_j^t = \mu^t + \zeta_j^t & \tilde{y}_j^c = \mu^c + 0 \\ \text{Within:} \quad \tilde{y}_{ij}^t = 0 + \varepsilon_{ij}^t & \tilde{y}_{ij}^c = 0 + \varepsilon_{ij}^c. \end{array} \quad (15)$$

Equations on the left represent the model for the treatment arm, and equations on the right represent the model for the control arm. Superscripts t or c refer to arms of the multiple-arm MSEM-PN within the model specification *only* and do not imply data management on the part of the researcher (i.e., the outcome is simply y_{ij} and does not need to be specially coded, as in the SSEM-PN). Although there are different ways to represent multiple-group (here, multiple-arm) models, here we chose to separately depict scalar equations implied by the MSEM-PN for each arm to most clearly highlight parameter equivalencies among MSEM-PN, SSEM-PN, and MLM-PN. The online Appendix provides the corresponding matrix representation for MSEM-PN, where parameter matrices differing across arm have a (g) superscript, and $g = t$ or c. Matrix representations are more conventional for MSEM (e.g., Kaplan, 2009; Liang & Bentler, 2004; B. O. Muthén & Asparouhov, 2009; Preacher, Zyphur, & Zhang, 2010).

In the treatment arm, the observed outcome in Equation (15) is decomposed into latent between ($\tilde{y}_j^t$) and latent within ($\tilde{y}_{ij}^t$) components. The tildes designate latent components. The latent between component of the treatment arm has mean μ^t and cluster specific deviation from the mean ζ_j^t, which is normally distributed:

$$\zeta_j^t \sim N(0, \psi^t). \quad (16)$$

The latent within component of the treatment arm, $\tilde{y}_{ij}^t$, has a 0 mean (a convention in MSEM is to estimate all means at the between level) and person-level residual ε_{ij}^t. In the control arm, Equation (15) decomposes y_{ij}^c into latent between ($\tilde{y}_j^c$) and within ($\tilde{y}_{ij}^c$) components, although ultimately variability is allowed in only one of the levels. At the between level, the mean μ^c of the latent between component in the control arm (corresponding to γ_{00} from MLM-PN) is estimated, but no cluster-specific deviation (ζ_j^c) is included. The average treatment effect is μ^t-μ^c (i.e., γ_{10} from MLM-PN). At the within level in the control arm, the latent component $\tilde{y}_{ij}^c$ has a 0 mean and person-level residual ε_{ij}^c. Person-level residuals in both arms are normally distributed and are independent by design:

$$\varepsilon_{ij}^c \sim N(0, \theta_\varepsilon^c) \quad \text{and} \quad \varepsilon_{ij}^t \sim N(0, \theta_\varepsilon^t). \quad (17)$$

Thus, in the treatment arm, the outcome variance is partitioned into ψ^t and θ_ε^t, as in SSEM-PN. ψ^t is equal to τ_{11} from the MLM-PN and θ_ε^t corresponds to σ_t^2 from MLM-PN. In the control arm, as in the SSEM-PN, an equivalent estimate of the outcome variance can be obtained *either* by estimating a variance for a person-level residual ε_{ij}^c (i.e., θ_ε^c) *or* by estimating a variance for a cluster-level residual ζ_j^c (i.e., ψ^c)—but not both. In Equations (15)–(17) we used the former option (which is illustrated in Figure 3, Panel A), although, for contrast, the latter option is depicted in Figure 3, Panel B, because we use it later. θ_ε^c corresponds to σ_c^2 from MLM-PN. Homoscedastic residual variances are obtained by constraining $\theta_\varepsilon^c = \theta_\varepsilon^t$, which then matches σ^2 from MLM-PN. Heteroscedastic residual variances are obtained by relaxing this constraint, as in Equation (17).

Panel A. MSEM-PN with typical identification method
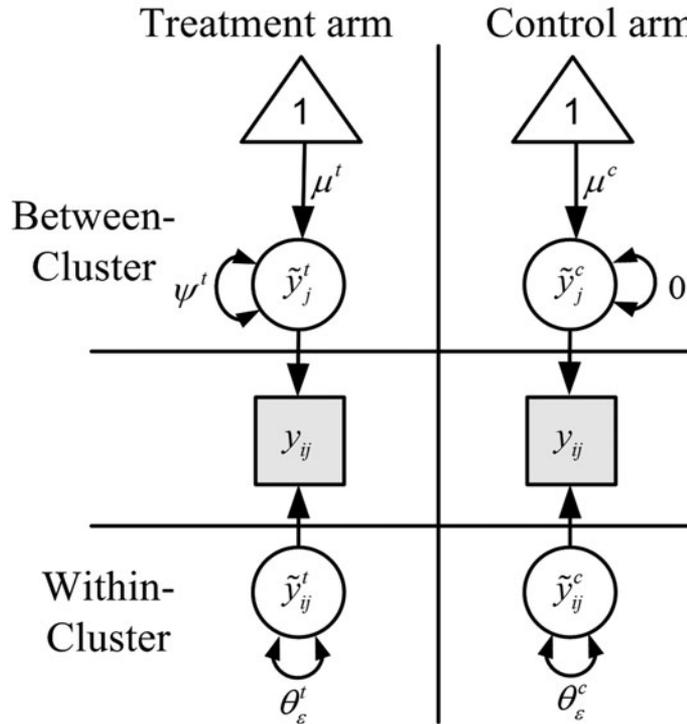
Panel B. MSEM-PN with alternative identification method
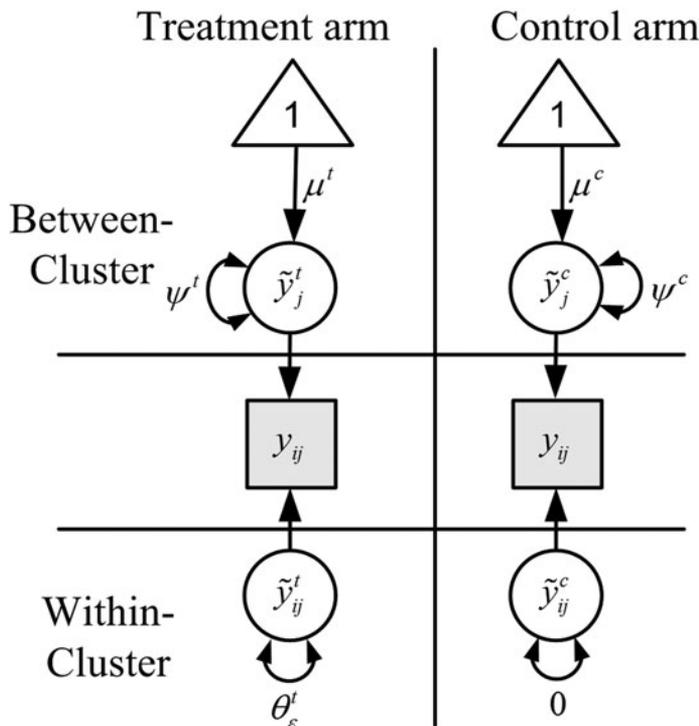
FIGURE 3   Multiple-arm multilevel structural equation model for partial nesting (MSEM-PN). *Note.* In each panel, the diagram is drawn for the ith person in the jth cluster in the treatment arm (left side) or control arm (right side). Notation in Panel A corresponds to the two-arm MSEM-PN with heteroscedastic residual variances in Equations (15)–(17). To simplify presentation, residuals ($\varepsilon_{ij}^t, \varepsilon_{ij}^c, \zeta_j^t, \zeta_j^c$) are not shown; only their variances are shown. Shapes are defined in the Figure 2 note.

This two-arm homoscedastic residual variance MSEM-PN is not saturated and contains no random slopes. Hence, its absolute fit could be examined as long as fit indices are computed using a modified saturated model incorporating the design-based constraints mentioned in the SSEM-PN section. Here, a modified saturated model is the two-arm heteroscedastic residual variance MSEM-PN from earlier. Generally, absolute fit indices for MSEM-PNs are more interesting to examine when additional theory-based constraints are imposed in the fitted model (see Discussion).

Adding Exogenous Covariates

Adding an exogenous person-level covariate x_{ij} and cluster-level covariate w_j to the MSEM-PN yields the following equations:

$$
\begin{array}{c|c}
\text{Treatment arm:} & \text{Control arm:} \\
y_{ij} = \tilde{y}_j^t + \tilde{y}_{ij}^t & y_{ij} = \tilde{y}_j^c + \tilde{y}_{ij}^c \\
\text{Between: } \tilde{y}_j^t = \mu^t + \omega_1^t w_j + \zeta_j^t & \tilde{y}_j^c = \mu^c + 0 \\
\text{Within: } \tilde{y}_{ij}^t = 0 + \kappa^t x_{ij} + \varepsilon_{ij}^t & \tilde{y}_{ij}^c = 0 + \kappa^c x_{ij} + \varepsilon_{ij}^c.
\end{array}
\tag{18}
$$

For controls, w_j is here missing-by-design, as is typical.[9] A single main effect of x_{ij} would be obtained by imposing the constraint $\kappa^t = \kappa^c$ to yield a single κ, which would match γ_{20} from the MLM-PN. Similar to SSEM-PN, allowing for an interaction between treatment and x_{ij} does not require formation of product terms; rather, this equality constraint must be relaxed. Finally, ω_1^t in Equation (18) corresponds with γ_{11} from the MLM-PN.

Adding Study Arms

A third arm may be added to the multiple-arm MSEM-PN to represent, say, a second treatment condition. Doing so requires decomposing variables in the third arm into their latent within and between components (where applicable) and specifying an arm-specific model for those components, as follows:

The conditional treatment effect in the third arm is μ^{t_2}-μ^c (γ_{30} in MLM-PN). In the third arm the total effect of x_{ij} is κ^{t_2}. Imposing the constraint $\kappa^c = \kappa^{t_1} = \kappa^{t_2}$ yields a single κ that matches γ_{20} in the MLM-PN. Relaxing this constraint again allows a treatment by x_{ij} interaction. If the third study arm is nonnested (e.g., if it is an individual therapy arm), w_j may be undefined, so we would impose $\omega_1^{t_2} = 0$ in Equation (19). Otherwise, the effect of w_j in the third arm is $\omega_1^{t_2}$ (γ_{31} in MLM-PN). Cluster-specific deviations are assumed independent across arm, due to the design, and are distributed as

$$
\zeta_j^{t_2} \sim N\left(0, \psi^{t_2}\right) \quad \text{and} \quad \zeta_j^{t_1} \sim N\left(0, \psi^{t_1}\right).
\tag{20}
$$

When the third arm is clustered, the variance, ψ^{t_2}, of cluster-specific deviation $\zeta_j^{t_2}$ corresponds with τ_{33} from the MLM-PN. Otherwise, if the third arm is nonnested, $\zeta_j^{t_2} = 0$ and $\psi^{t_2} = 0$. Residuals, also independent across arm, are distributed as follows for heteroscedastic residual variances across arm:

$$
\varepsilon_{ij}^c \sim N(0, \theta_\varepsilon^c)
$$

$$
\text{and} \quad \varepsilon_{ij}^{t_1} \sim N(0, \theta_\varepsilon^{t_1})
$$

$$
\text{and} \quad \varepsilon_{ij}^{t_2} \sim N(0, \theta_\varepsilon^{t_2}).
\tag{21}
$$

$\theta_\varepsilon^{t_2}$ matches $\sigma_{t_2}^2$ in MLM-PN. The equality constraint $\theta_\varepsilon^c = \theta_\varepsilon^{t_1} = \theta_\varepsilon^{t_2}$ is imposed in MSEM-PN to obtain homoscedastic[4] residual variances (matching σ^2 in MLM-PN).

SIMULATED EXAMPLES

We now use simulated examples to demonstrate that our two SEM approaches for partial nesting (SSEM-PN and MSEM-PN) equivalently implement each basic model specification previously considered for MLM-PN. A single sample was generated from each of four different MLM-PNs. Table 1 lists generating parameters for two-arm homoscedastic versus two-arm heteroscedastic

$$
\begin{array}{c|c|c}
\text{Treatment 1 arm:} & \text{Treatment 2 arm:} & \text{Control arm:} \\
y_{ij} = \tilde{y}_j^{t_1} + \tilde{y}_{ij}^{t_1} & y_{ij} = \tilde{y}_j^{t_2} + \tilde{y}_{ij}^{t_2} & y_{ij} = \tilde{y}_j^c + \tilde{y}_{ij}^c \\
\text{Between: } \tilde{y}_j^{t_1} = \mu^{t_1} + \omega_1^{t_1} w_j + \zeta_j^{t_1} & \tilde{y}_j^{t_2} = \mu^{t_2} + \omega_1^{t_2} w_j + \zeta_j^{t_2} & \tilde{y}_j^c = \mu^c + 0 \\
\text{Within: } \tilde{y}_{ij}^{t_1} = 0 + \kappa^{t_1} x_{ij} + \varepsilon_{ij}^{t_1} & \tilde{y}_{ij}^{t_2} = 0 + \kappa^{t_2} x_{ij} + \varepsilon_{ij}^{t_2} & \tilde{y}_{ij}^c = 0 + \kappa^c x_{ij} + \varepsilon_{ij}^c.
\end{array}
\tag{19}
$$

[9]As described in the MLM-PN section (Bauer et al., 2008) assigning control cases an arbitrary value that is not the missing data code (e.g., 0) for missing-by-design scores on w_j prevents listwise deletion of these cases when using FIML with exogenous predictors and a conditional likelihood. Also, MSEM-PN requires w_j to have nonzero variance in each arm; this can be addressed in Mplus by specifying *variances = nocheck* or more generally by assigning at least two different arbitrary values for missing-by-design scores on w_j. Similar to MLM-PN, the choice of what arbitrary values to assign does not affect results (as the slope of w_j is fixed to 0 for controls; see also Bauer et al., 2008).

residual variance MLM-PNs. Table 2 lists generating parameters for two-arm versus three-arm homoscedastic residual variance MLM-PNs with a Level 1 covariate, x_{ij}, in all arms, and Level 2 covariate, w_j, in the treatment arm(s). The pattern of generating parameter values is consistent with what might be encountered in a treatment-outcome study on a problem behavior (e.g., internalizing) in which (a) individuals are randomly assigned to condition, (b) treatment on average led to a decrease in the problem behavior (although the effect was not

TABLE 1
Two-Arm Partial Nesting Model With Homoscedastic or Heteroscedastic Across-Arm Residual Variances:
Comparison of Alternative MLM-PN, SSEM-PN, and MSEM-PN Specifications for Simulated Single-Sample Data Sets

	Notation			Results of Model Fitting						
				2-Arm Homoscedastic Residual Variance				2-Arm Heteroscedastic Residual Variance		
Parameter Description	MLM-PN Eqn 1–3	SSEM-PN Eqn 7–9 & MSEM-PN Eqn 15–17	Pop-Gen Val.	MLM-PN Est (SE)	SSEM-PN Est (SE)	MSEM-PN Est (SE)	Pop-Gen Val.	MLM-PN Est (SE)	SSEM-PN Est (SE)	MSEM-PN Est (SE)
Mean, cont arm	γ_{00}	μ^c	4	4.00 (.10)	4.00 (.10)	4.00 (.10)	4	3.92 (.10)	3.92 (.10)	3.92 (.10)
Mean, tx arm†	$\gamma_{00} + \gamma_{10}$	μ^t	1	1.03 (.17)	1.03 (.17)	1.03 (.17)	1	.85 (.16)	.85 (.17)	.85 (.17)
Tx effect†	γ_{10}	$\mu^t - \mu^c$	−3	−2.97 (.20)	−2.97 (.20)	−2.97 (.20)	−3	−3.07 (.19)	−3.07 (.19)	−3.07 (.19)
Lev 2 var, tx arm	τ_{11}	ψ^t	.75	.73 (.26)	.73 (.26)	.73 (.26)	.75	.77 (.25)	.77 (.25)	.77 (.25)
Var, cont arm*	σ_c^2	θ_ε^c	2	2.02 (.15)	2.02 (.15)	2.02 (.15)	2	2.09 (.21)	2.09 (.21)	2.09 (.21)
Lev 1 var, tx arm*	σ_t^2	θ_ε^t	2	2.02 (.15)	2.02 (.15)	2.02 (.15)	1.5	1.59 (.18)	1.59 (.18)	1.59 (.18)
Log-likelihood				−728.65	−728.65	−728.65		−712.27	−712.27	−712.27

Note. †Compound coefficients (sums or differences of estimated parameters) had *SE*s computed via the delta method using the model constraint command in Mplus or using the estimate statement in SAS Proc Mixed. *Homoscedastic residual variances require constraining these variances equal. Notation was defined in the text. MLM-PN = multilevel model for partial nesting; SSEM-PN = multivariate single-level structural equation model (SEM) for partial nesting; MSEM-PN = multiple-arm multilevel SEM for partial nesting; var = variance; Eqn = equation in the text; val. = value; lev = level; tx = treatment; cont = control; est = estimate; gen = generating; Pop = population.

the same for both treatments), and (c) outcomes varied across clusters in all arms and within cluster in treatment arm(s). In the two-arm examples, the treatment arm had 40 clusters of 5 persons each; the control arm had 200 persons (i.e., 200 clusters of size 1). Hence, $N = 400$, $J = 240$, and $N_j = 1$ or 5. In the three-arm example, the second treatment arm contained 100 clusters of size 2 (i.e., 200 persons). Example data were

generated in SAS 9.2 such that, marginally, approximately one third of the variability in y_{ij} was at the between level in a treatment arm ($ICC_y = .27$ and .33 for the two models in Table 1; $ICC_y = .29$ and .33 for the two models in Table 2). There are a number of reasons a third of the outcome variation in the treatment arm(s) would be between clusters. These reasons include differing fidelity/effectiveness of treatment

TABLE 2
Two-Arm or Three-Arm Partial Nesting Model With Covariates and Homoscedastic Residual Variances:
Comparison of Alternative MLM-PN, SSEM-PN, and MSEM-PN Specifications With Simulated Single-Sample Data Sets

	Notation			Results of Model Fitting						
				2-Arm With Covariates				3-Arm With Covariates		
Parameter Description	MLM-PN Eqn 4–6	SSEM-PN Eqn 10–14 & MSEM-PN Eqn 18–21	Pop-Gen Val.	MLM-PN Est (SE)	SSEM-PN Est (SE)	MSEM-PN Est (SE)	Pop-Gen Val.	MLM-PN Est (SE)	SSEM-PN Est (SE)	MSEM-PN Est (SE)
Intercept cont arm	γ_{00}	μ^c	4	4.02 (.10)	4.02 (.10)	4.02 (.10)	4	3.95 (.10)	3.95 (.10)	3.95 (.10)
Intercept tx1 arm†	$\gamma_{00} + \gamma_{10}$	μ^{t_1}	1	.98 (.15)	.98 (.16)	.98 (.16)	1	1.04 (.17)	1.04 (.17)	1.04 (.17)
Condit. tx1 effect†	γ_{10}	$\mu^{t_1} - \mu^c$	−3	−3.05 (.19)	−3.05 (.19)	−3.05 (.19)	−3	−2.91 (.20)	−2.91 (.20)	−2.91 (.20)
Intercept tx2 arm†	$\gamma_{00} + \gamma_{30}$	μ^{t_2}					2	1.92 (.12)	1.92 (.12)	1.92 (.12)
Condit. tx2 effect†	γ_{30}	$\mu^{t_2} - \mu^c$					−2	−2.03 (.15)	−2.03 (.15)	−2.03 (.15)
L2 res. var, tx1 arm	τ_{11}	ψ^{t_1}	.75	.57 (.22)	.57 (.22)	.57 (.22)	.75	.81 (.26)	.81 (.26)	.81 (.26)
L2 res. var, tx2 arm	τ_{33}	ψ^{t_2}					.50	.46 (.20)	.46 (.20)	.46 (.20)
Res. var, cont arm*	σ_c^2	θ_ε^c	2	2.10 (.16)	2.10 (.16)	2.10 (.16)	2	1.85 (.12)	1.85 (.12)	1.85 (.12)
L1 res. var, tx1 arm*	$\sigma_{t_1}^2$	$\theta_\varepsilon^{t_1}$	2	2.10 (.16)	2.10 (.16)	2.10 (.16)	2	1.85 (.12)	1.85 (.12)	1.85 (.12)
L1 res. var, tx2 arm*	$\sigma_{t_2}^2$	$\theta_\varepsilon^{t_2}$					2	1.85 (.12)	1.85 (.12)	1.85 (.12)
Effect x_{ij}	γ_{20}	$\kappa^c = \kappa^{t_1} = \kappa^{t_2}$	1	1.09 (.08)	1.09 (.08)	1.09 (.08)	1	1.06 (.06)	1.06 (.06)	1.06 (.06)
Effect w_j, tx1 arm	γ_{11}	$\omega_1^{t_1}$	.50	.43 (.15)	.43 (.15)	.43 (.15)	.50	.30 (.17)	.30 (.17)	.30 (.17)
Effect w_j, tx2 arm	γ_{31}	$\omega_1^{t_2}$					.50	.47 (.11)	.47 (.11)	.47 (.11)
Log-likelihood				−733.42	−733.42	−733.42		−1079.44	−1079.44	−1079.44

Note. The 3-arm generating and fitted models in Table 2 had clustering in both treatment arms (abbreviated tx1 and tx2); results for a 3-arm study with an unclustered treatment arm are not shown. *Generating and fitted models in Table 2 had homoscedastic residual variances across arms, which was accomplished by constraining these variances equal. †Compound coefficients (sums or differences of estimated parameters) had *SE*s computed via the delta method using a model constraint in Mplus or estimate statement in SAS Mixed. The effect of *x* controls for *w* (and vice versa) in treatment arm(s). Notation is defined in the text. Condit = conditional. See Table 1 note for abbreviation definitions. Res. = Residual; L2 = Level 2; L1 = Level 1.

implementation across therapy clusters. When present, w_j and x_{ij} were uncorrelated and standard normally distributed. One third of the variability in x_{ij} was at the between level in the treatment arm(s) (i.e., $ICC_x = .33$). MLM-PNs were fit in SAS 9.2[10] and SSEM-PN and MSEM-PN were fit in Mplus 7, all using FIML.

Table 1 summarizes corresponding MLM-PN, SSEM-PN, and MSEM-PN notation for the two-arm homoscedastic versus heteroscedastic residual variance models alongside matching results from fitting these three models to the simulated example data. Table 2 provides a summary of corresponding notation for MLM-PN, SSEM-PN, and MSEM-PN for two-arm versus three-arm homoscedastic residual variance models with covariates and shows all three specifications provide matching results.

Summary

Using equations and diagrams, we have described two SEM approaches for partial nesting (SSEM-PN and MSEM-PN). Using simulation, we showed that they equivalently implement every basic MLM-PN specification. We next turn to two extensions handled better by SEM approaches than MLM-PN. The first extension concerns multivariate outcomes, some of which exist at the cluster level. The second extension involves decomposing effects of person-level covariates.

CLUSTER-LEVEL OUTCOMES: CONTRIBUTION OF SSEM-PN AND MSEM-PN

Treatment studies often involve multivariate outcomes, some of which may be measured at the cluster level (say, therapy-group cooperation or implementation fidelity). In MLM-PN (and MLM generally) incorporating Level 1 (here, person-level) multivariate outcomes usually requires some data management to trick[11] univariate MLM software into fitting the

---

[10]MLM-PNs were rerun using alternative methods for computing degrees of freedom for t tests of fixed effects (the Kenward-Rogers method from Baldwin et al. [2011] and Bauer et al., [2008], which is available in only some multilevel modeling packages, and a Containment method). Standard errors were the same to three decimal places. In contrast, SEM packages typically provide only z tests for fixed effects. Baldwin et al. showed that, unless there were very few clusters in the treatment arm (<8), alternative df computation methods provided equally good Type I error rates in MLM-PN. Our empirical example and simulations had more clusters than this in the treatment arm.

[11]In this MLM trick, multivariate outcomes are concatenated vertically in the data set, and one toggle indicator variable is constructed per outcome (e.g., the first toggle indicator is 1 if y_{ij} refers to the first outcome, 0 otherwise). A univariate MLM is fit, but no overall intercept is estimated; rather, main effects of the indicator variables refer to outcome-specific intercepts. Product terms of the indicator variables with each predictor are entered to allow the predictors to have outcome-specific effects.

model (Bauer, Preacher, & Gil, 2006; Compas et al., 2009; Goldstein, 1995; Leyland & Goldstein, 2001). Furthermore, MLM-PN (and MLM generally) cannot include *cluster-level* outcomes among these multivariate outcomes (e.g., Croon & van Veldhoven, 2007). In contrast, a cluster-level outcome is straightforward to add in SSEM-PN and MSEM-PN, as described later. We assume below that the cluster-level outcome is measured only in the treatment arm and is missing-by-design in the control arm; however, this need not be the case. For instance, the outcome "problem-solving speed" could be measured for teams in the treatment arm but for persons in the control arm—recalling that each control arm person can be considered a cluster (e.g., Figure 2, Panel B; Figure 3, Panel B).

Equation (22) incorporates a cluster-level outcome z_j into the treatment arm of the two-arm SSEM-PN from Equation (10b). In this formulation, parameters pertaining to the original outcome are superscripted y and parameters pertaining to the new outcome are superscripted z. In the treatment arm, μ^{zt} is the conditional mean of z_j and ω_1^{zt} is the slope of z_j on w_j. Here z_j can have a residual only at the cluster level, ζ_j^{zt} (so its person-level residual variance is fixed to 0).

$$
\begin{bmatrix} y_{1j}^c \\ y_{1j}^t \\ y_{2j}^t \\ y_{3j}^t \\ y_{4j}^t \\ y_{5j}^t \\ z_j^t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_j^{yc} \\ \eta_j^{yt} \\ \eta_j^{zt} \end{bmatrix}
$$

$$
+ \begin{bmatrix} \kappa^{yc} & 0 & 0 & 0 & 0 \\ \kappa^{yt} & 0 & 0 & 0 & 0 \\ 0 & \kappa^{yt} & 0 & 0 & 0 \\ 0 & 0 & \kappa^{yt} & 0 & 0 \\ 0 & 0 & 0 & \kappa^{yt} & 0 \\ 0 & 0 & 0 & 0 & \kappa^{yt} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ x_{4j} \\ x_{5j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1j}^{yc} \\ \varepsilon_{1j}^{yt} \\ \varepsilon_{2j}^{yt} \\ \varepsilon_{3j}^{yt} \\ \varepsilon_{4j}^{yt} \\ \varepsilon_{5j}^{yt} \\ 0 \end{bmatrix}
$$

(22)

$$
\begin{bmatrix} \eta_j^{yc} \\ \eta_j^{yt} \\ \eta_j^{zt} \end{bmatrix} = \begin{bmatrix} \mu^{yc} \\ \mu^{yt} \\ \mu^{zt} \end{bmatrix} + \begin{bmatrix} 0 \\ \omega_1^{yt} \\ \omega_1^{zt} \end{bmatrix} \begin{bmatrix} w_j \end{bmatrix} + \begin{bmatrix} 0 \\ \zeta_j^{yt} \\ \zeta_j^{zt} \end{bmatrix}.
$$

In the treatment arm, ζ_j^{zt} can covary with cluster-level residuals for the original outcome, ζ_j^{yt}:

$$
\begin{bmatrix} 0 \\ \zeta_j^{yt} \\ \zeta_j^{zt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & & \\ 0 & \psi^{yt} & \\ 0 & \psi^{yzt} & \psi^{zt} \end{bmatrix} \right).
$$

(23)

Next, Equation (24) incorporates a cluster-level outcome z_j into the treatment arm of the two-arm MSEM-PN

from Equation (18); parameters pertaining to each outcome are superscripted y or z accordingly. z_j has only a latent between component, $\tilde{z}_j^t$, because it lacks within-cluster variability.

$$
\begin{array}{l}
\quad\quad\quad \text{Treatment arm:} \quad\quad\quad\quad \text{Control arm:} \\
\quad\quad\quad y_{ij} = \tilde{y}_j^t + \tilde{y}_{ij}^t \quad\quad\quad\quad\quad y_{ij} = \tilde{y}_j^c + \tilde{y}_{ij}^c \\
\quad\quad\quad z_j = \tilde{z}_j^t \\
\text{Between:} \quad \tilde{y}_j^t = \mu^{yt} + \omega_1^{yt} w_j + \zeta_j^{yt} \quad\quad \tilde{y}_j = \mu^{yc} + 0 \\
\quad\quad\quad \tilde{z}_j^t = \mu^{zt} + \omega_1^{zt} w_j + \zeta_j^{zt} \\
\text{Within:} \quad\quad \tilde{y}_{ij}^t = 0 + \kappa^{yt} x_{ij} + \varepsilon_{ij}^{yt} \quad\quad \tilde{y}_{ij}^c = 0 + \kappa^{yc} x_{ij} + \varepsilon_{ij}^{yc}.
\end{array}
\tag{24}
$$

μ^{zt} and ω_1^{zt} have the same interpretation as in the corresponding SSEM-PN (Equation (22)). ζ_j^{zt} and ζ_j^{yt} are distributed as in Equation (23), and ε_{ij}^{yt} and ε_{ij}^{yc} are distributed as in Equation (17). The empirical example presented later will include a cluster-level outcome in the treatment arm to illustrate further this addition to the SSEM-PN and MSEM-PN.

DECOMPOSING EFFECTS OF PERSON-LEVEL COVARIATES IN THE TREATMENT ARM: CONTRIBUTION OF SSEM-PN AND MSEM-PN

Options for including a person-level covariate x_{ij} in partial nesting designs have not been thoroughly discussed in the literature. To accomplish this, we must distinguish among three possible effects of x_{ij} in the treatment arm: the total (conflated) effect of x_{ij} versus within effect of x_{ij} versus between effect of x_{ij}. The *within effect* for persons in cluster j represents the change in y_{ij} per unit increase in x_{ij} controlling for the cluster j average of x_{ij}. The *between effect* represents the change in y_{ij} per unit increase in the cluster j average. Including only an uncentered (or grand-mean centered) x_{ij}, as has been done in prior published MLM-PN data analyses—and earlier in this article—yields a total (conflated) effect in the treatment arm, which is a weighted average of within and between effects. Researchers, however, have been increasingly interested in decomposing a total effect of x_{ij} into within and between effects (e.g., Curran & Bauer, 2011; Curran, Lee, Howard, Lane, & MacCallum, 2012; Enders & Tofighi, 2007; Hedeker & Gibbons 2006; Hofmann & Gavin, 1998; Kreft, de Leeuw, & Aiken, 1995; Neuhaus & Kalbfleisch, 1998).

First we discuss an *exogenous observed cluster mean approach* for decomposing the effect of a person-level covariate x_{ij} in the treatment arm of a partially nested design—using two strategies adapted from the MLM literature. These strategies are termed Strategy A and Strategy B. Here we implement Strategy A with MLM-PN and MSEM-PN and Strategy B with MLM-PN and SSEM-PN. We show that when

MSEM-PN[12] (for Strategy A) or SSEM-PN[13] (for Strategy B) employ the exogenous observed cluster mean approach, they replicate MLM-PN results, but arguably with a simpler specification.

Next, we discuss an alternative *endogenous latent cluster mean approach* for decomposing the effect of a person-level covariate x_{ij} in the treatment arm of a partially nested design; this approach is available only for MSEM-PN and SSEM-PN. This approach has an advantage over the exogenous observed cluster mean approach available for MLM-PN because the latent cluster mean of x is more reliable. Hence, its effects are less subject to attenuation (e.g., Asparouhov & Muthén, 2007; Curran et al., 2012; Lüdtke et al., 2008; Preacher et al., 2010). Here we implement the endogenous latent cluster mean approach for MSEM-PN and SSEM-PN. Following the presentation of both approaches for decomposing the effect of x_{ij} in the treatment arm, both are demonstrated using a simulated partially nested example in which between and within effects of x_{ij} are unequal in the treatment arm.

Exogenous Observed Cluster Mean Approach: Strategy A

In MLM, Strategy A involves including the cluster-mean-centered $\dot{x}_{ij}$ (where $\dot{x}_{ij} = x_{ij} - \bar{x}_{.j}$) as an exogenous person-level predictor and the observed cluster mean $\bar{x}_{.j}$ (where $\bar{x}_{.j} = \sum_{i=1}^{N_j} x_{ij} \big/ N_j$) as an exogenous cluster-level predictor (e.g., Bauer et al., 2008; Raudenbush & Bryk, 2002). In MLM, the slope of $\dot{x}_{ij}$ yields a within effect and the slope of $\bar{x}_{.j}$ yields a between effect. This strategy can be adapted to MLM-PN by including $\dot{x}_{ij}$ and $\bar{x}_{.j} treat_{ij}$ as predictors, so $\bar{x}_{.j}$ explains only between-cluster variability in the treatment arm. Yet, this strategy has a limitation in the partial nesting context. It precludes $\dot{x}_{ij}$ from explaining variability in the control arm because $\dot{x}_{ij}$ is 0 for controls after cluster-mean centering. To overcome this limitation and estimate the predictor's effect for controls, $x_{ij}(1 - treat_{ij})$ is also included. To summarize Strategy A for MLM-PN,

$$
\begin{aligned}
\text{Level 1: } y_{ij} &= \beta_{0j} + \beta_{1j} treat_{ij} + \beta_{2j} \dot{x}_{ij} \\
&\quad + \beta_{3j} x_{ij}(1 - treat_{ij}) + r_{ij} \\
\text{Level 2: } \beta_{0j} &= \gamma_{00} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11} w_j + \gamma_{12} \bar{x}_{.j} + u_{1j} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30}.
\end{aligned}
\tag{25}
$$

---

[12]Strategy B also applies to MSEM-PN; it is less commonly implemented, so is omitted to save space.

[13]Strategy A is not estimable in the SSEM-PN due to the ipsativity described in Curran et al. (2012).

Reduced form:

$$y_{ij} = \gamma_{00} + (\gamma_{10} + \gamma_{11}w_j + \gamma_{12}\bar{x}_{.j} + u_{1j})treat_{ij}$$
$$+ \gamma_{20}\dot{x}_{ij} + \gamma_{30}x_{ij}(1 - treat_{ij}) + r_{ij}.$$

Here, γ_{20} is the within effect of x_{ij} for treated persons, γ_{12} is the between effect for treated persons, and γ_{30} is the effect for control persons. To generalize to three study arms, two of which are clustered, would require first replacing $x_{ij}(1 - treat_{ij})$ with $x_{ij}(1 - treat1_{ij} - treat2_{ij})$ to obtain the effect of x_{ij} in the control arm. Also, separate within and between effects of x_{ij} in each treatment arm would need to be represented by four product terms in the reduced form model (i.e., using $\dot{x}_{ij}treat2_{ij}$ and $\dot{x}_{ij}treat1_{ij}$ as predictors instead of $\dot{x}_{ij}$, and using $\bar{x}_{.j}treat1_{ij}$ and $\bar{x}_{.j}treat2_{ij}$ as predictors instead of $\bar{x}_{.j}treat_{ij}$).

For MSEM-PN, the following changes are made to the treatment arm of Equation (18) to implement Strategy A. Exogenous x_{ij} is replaced by $\dot{x}_{ij}$ in the within model (its coefficient, κ_W^t, now has a W subscript to denote it is the within effect of x_{ij}), and the observed $\bar{x}_{.j}$ is added to the between model (its coefficient, ω_{2B}^t, has a B subscript to emphasize that it is the between effect of x_{ij}):

	Treatment arm:	Control arm:
	$y_{ij} = \bar{y}_j^t + \bar{y}_{ij}^t$	$y_{ij} = \bar{y}_j^c + \bar{y}_{ij}^c$
Between:	$\bar{y}_j^t = \mu^t + \omega_1^t w_j + \omega_{2B}^t \bar{x}_{.j} + \zeta_j^t$	$\bar{y}_j^c = \mu^c + 0$
Within:	$\bar{y}_{ij}^t = 0 + \kappa_W^t \dot{x}_{ij} + \varepsilon_{ij}^t$	$\bar{y}_{ij}^c = 0 + \kappa^c x_{ij} + \varepsilon_{ij}^c.$

(26)

This specification is diagrammed in Figure 4, Panel A. The extension to obtain separate between and within effects of x_{ij} in two clustered treatment arms of a three-arm study requires the same changes in each treatment arm of Equation (19), yielding $\kappa_W^{t_1}$, $\omega_{2B}^{t_1}$, $\kappa_W^{t_2}$, and $\omega_{2B}^{t_2}$.

Exogenous Observed Cluster Mean Approach: Strategy B

In MLM, Strategy B for decomposing effects across levels involves including the exogenous *raw* x_{ij} as well as $\bar{x}_{.j}$ as predictors. The former yields a within effect; the latter yields a *contextual effect* (Raudenbush & Bryk, 2002). A contextual effect—defined as the between effect minus the within effect—can then be used straightforwardly to solve for the between effect. (The between effect is the sum of within and contextual effects.) This strategy is adapted for a two-arm MLM-PN by including $x_{ij}treat_{ij}$, $x_{ij}(1 - treat_{ij})$, and $\bar{x}_{.j}treat_{ij}$ as predictors in the reduced form of Equation (27).

$$\text{Level 1: } y_{ij} = \beta_{0j} + \beta_{1j}treat_{ij} + \beta_{2j}x_{ij}treat_{ij}$$
$$+ \beta_{3j}x_{ij}(1 - treat_{ij}) + r_{ij} \qquad (27)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_j + \gamma_{12}\bar{x}_{.j} + u_{1j}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$

Reduced form:

$$y_{ij} = \gamma_{00} + (\gamma_{10} + \gamma_{11}w_j + \gamma_{12}\bar{x}_{.j} + u_{1j})treat_{ij}$$
$$+ \gamma_{20}x_{ij}treat_{ij} + \gamma_{30}x_{ij}(1 - treat_{ij}) + r_{ij}.$$

In the control arm, γ_{30} is the effect of x_{ij}. In the treatment arm, γ_{20} is the within effect of x_{ij}; γ_{12} is the contextual effect of x_{ij}. The between effect of x_{ij} is $\gamma_{12} + \gamma_{20}$. Obtaining between and within effects of x_{ij} in two clustered treatment arms of a *three-arm* study requires the following replacements in the reduced form expression: We replace $x_{ij}treat_{ij}$ by $x_{ij}treat1_{ij}$ and $x_{ij}treat2_{ij}$; we also replace $x_{ij}(1 - treat_{ij})$ by $x_{ij}(1 - treat1_{ij} - treat2_{ij})$ and replace $\bar{x}_{.j}treat_{ij}$ by $\bar{x}_{.j}treat1_{ij}$ and $\bar{x}_{.j}treat2_{ij}$.

Strategy B is straightforward to implement in SSEM-PN, as follows. Using the SSEM-PN data set organization described previously, rank-deficiency problems do not emerge.[14] First, we add $\bar{x}_{.j}$ as an exogenous predictor in the treatment arm in the structural model of Equation (10b), yielding

$$\begin{bmatrix} \eta_j^c \\ \eta_j^t \end{bmatrix} = \begin{bmatrix} \mu^c \\ \mu^t \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \omega_1^t & \omega_2^t \end{bmatrix} \begin{bmatrix} w_j \\ \bar{x}_{.j} \end{bmatrix} + \begin{bmatrix} 0 \\ \zeta_j^t \end{bmatrix}. \qquad (28)$$

ω_2^t is the contextual effect in the treatment arm. Also, in Equation (10b), κ^t can now be relabeled κ_W^t because it is now the within effect of x_{ij} in the treatment arm. Adding $\omega_2^t + \kappa_W^t$ yields the between effect. This specification is diagrammed in Figure 5, Panel A. For a three-arm study with two clustered treatment arms, $\bar{x}_{.j}$ similarly can be included in Equation (12) as an exogenous predictor in both treatment arms; $\omega_2^{t_1}$ and $\omega_2^{t_2}$ would be contextual effects in each treatment arm. Then κ^{t_1} and κ^{t_2} in Equation (12) could be relabeled $\kappa_W^{t_1}$ and $\kappa_W^{t_2}$ to underscore that they are now within effects in each treatment arm.

Endogenous Latent Cluster Mean Approach

Both MSEM-PN and SSEM-PN can employ an attractive alternative approach to decompose effects of person-level predictors in the treatment arm, involving substituting $\bar{x}_{.j}$ (which is subject to error) for an endogenous latent cluster mean of x (i.e., random intercept of x). We discuss implementing this approach for MSEM-PN and then SSEM-PN.

---

[14]If a predictor (i.e., column in a wide-format data set) is an exact linear function of other predictors (columns) the covariance matrix of predictors has a zero determinant. However, in the partial nesting data structure, $\bar{x}_{.j}$ is not the same linear function of $x_{1j} - x_{5j}$ for all rows of the data set. $\bar{x}_{.j}$ is calculated from the present data only and so will be a different function depending on how many people are in a given cluster. In the control arm, only x_{1j} is used in computing $\bar{x}_{.j}$. As mentioned previously, in the control arm, $y_{2j} - y_{5j}$ and $x_{2j} - x_{5j}$ are missing-by-design and the latter are assigned an arbitrary numeric value (not the missing code) to retain the control persons in the conditional likelihood computation (see Footnote 8 and see Bauer et al., 2008).
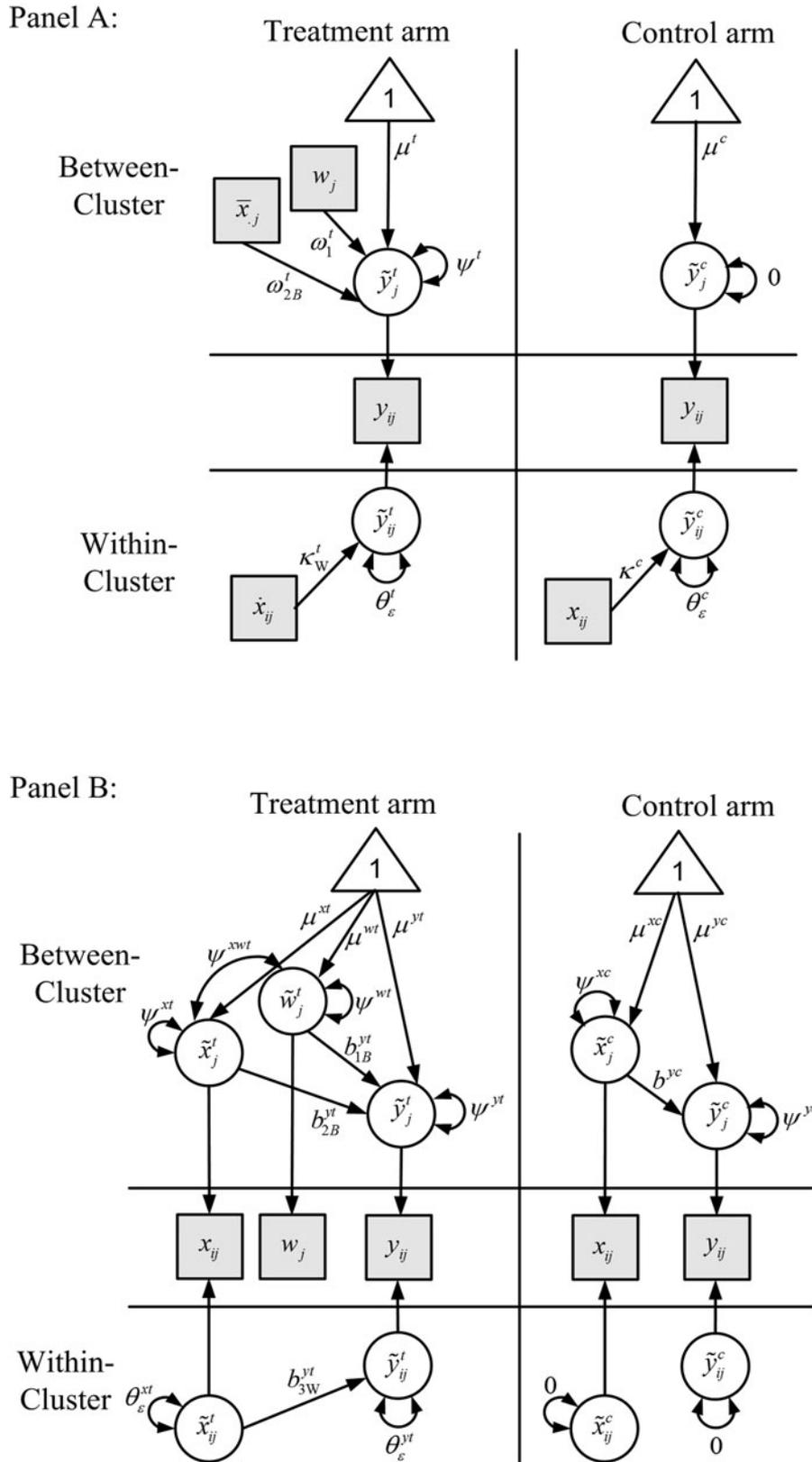
FIGURE 4   MSEM-PN that decomposes a person-level predictor's (x_{ij}) effect in the treatment arm using an exogenous observed cluster mean (Panel A) versus an endogenous latent cluster mean (Panel B). *Note.* All predictors are exogenous in Panel A and endogenous in Panel B. Panel A uses Strategy A from the text. Notation for Panel A matches Equation (26) where $\theta_\varepsilon^t = \theta_\varepsilon^c$ and for Panel B matches Equations (29)–(31) where $\theta_\varepsilon^{yt} = \psi^{yc}$. To simplify presentation, residuals are not shown; only their variances are shown. MSEM-PN = multiple-arm multilevel structural equation model for partial nesting.
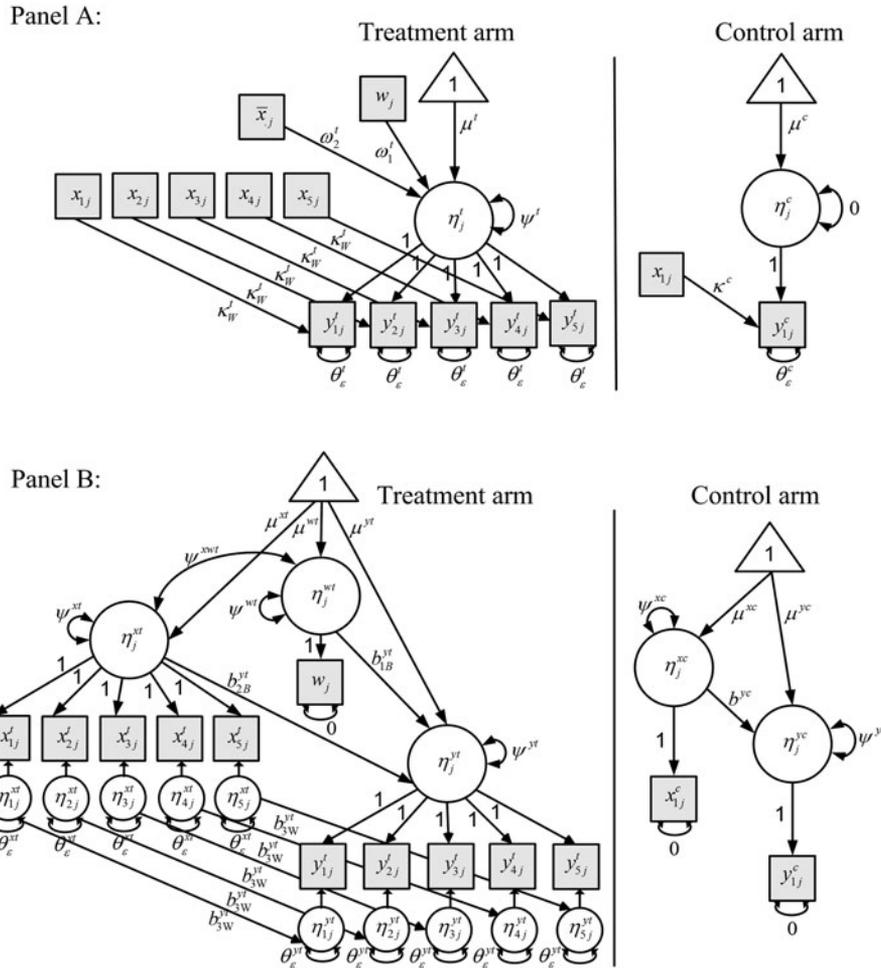
Panel A:

Panel B:

FIGURE 5   SSEM-PN that decomposes a person-level predictor's (x_{ij}) effect in the treatment arm using an exogenous observed cluster mean (Panel A) versus an endogenous latent cluster mean (Panel B). *Note.* All predictors are exogenous in Panel A and endogenous in Panel B. Panel A uses Strategy B from the text. Panel A notation corresponds with the SSEM-PN in Equation (28) and accompanying text, when $\theta_\varepsilon^t = \theta_\varepsilon^c$. Panel B notation corresponds with the SSEM-PN in Equations (32)–(34) when $\theta_\varepsilon^{yt} = \psi^{yc}$. To simplify presentation, residuals are not shown; only their variances are shown. SSEM-PN = multivariate single-level structural equation model for partial nesting.

For MSEM-PN, this approach requires replacing the exogenous observed components of x_{ij} (i.e., $\dot{x}_{ij}$ and $\bar{x}_{.j}$ from Equation (26)) with the endogenous latent within and between components of x_{ij} (denoted $\tilde{x}_{ij}^t$ and $\tilde{x}_j^t$) in the treatment arm of Equation (29).

This specification is diagrammed in Figure 4, Panel B. $\tilde{x}_j^t$ is called a latent cluster mean (or random intercept) and is more reliable than $\bar{x}_{.j}$ (Lüdtke et al., 2008). In Equation (29), bs represent effects of latent components. In the treatment arm, within and between effects of x_{ij} are b_{3W}^{yt} and b_{2B}^{yt},

$$
\begin{array}{ll}
\text{Treatment arm:} & \text{Control arm:} \\
y_{ij} = \tilde{y}_j^t + \tilde{y}_{ij}^t & y_{ij} = \tilde{y}_j^c + \tilde{y}_{ij}^c \\
x_{ij} = \tilde{x}_j^t + \tilde{x}_{ij}^t & x_{ij} = \tilde{x}_j^c + \tilde{x}_{ij}^c \\
w_j^t = \tilde{w}_j^t & \\
\text{Between:} \quad \tilde{y}_j^t = \mu^{yt} + b_{1B}^{yt}\tilde{w}_j^t + b_{2B}^{yt}\tilde{x}_j^t + \zeta_j^{yt} & \tilde{y}_j^c = \mu^{yc} + b^{yc}\tilde{x}_j^c + \zeta_j^{yc} \\
\tilde{x}_j^t = \mu^{xt} + \zeta_j^{xt} & \tilde{x}_j^c = \mu^{xc} + \zeta_j^{xc} \\
\tilde{w}_j^t = \mu^{wt} + \zeta_j^{wt} & \\
\text{Within:} \quad \tilde{y}_{ij}^t = 0 + b_{3W}^{yt}\tilde{x}_{ij}^t + \varepsilon_{ij}^{yt} & \tilde{y}_{ij}^c = 0 \\
\tilde{x}_{ij}^t = 0 + \varepsilon_{ij}^{xt} & \tilde{x}_{ij}^c = 0.
\end{array}
\tag{29}
$$

respectively. Because w_j varies only at the cluster level in the treatment arm, it has only a latent between component, $\tilde{w}_j^t$. In Equation (29), the between effect of w_j is b_{1B}^{yt}. Of the latent components for x_{ij} in the control arm ($\tilde{x}_{ij}^c$ and $\tilde{x}_j^c$), *only one* component is allowed to have variation, given the lack of nesting in that arm. Here we chose $\tilde{x}_j^c$, corresponding to the cluster level. To streamline presentation, we also allowed outcome variation only for the cluster-level component in the control arm, $\tilde{y}_j^c$, as previously illustrated in Figure 3, Panel B. Hence, residual variances are here estimated only at the cluster level in the control arm (Equation (30)). But in the treatment arm, residual variances are estimated at cluster and person levels (Equations (30)–(31)).

$$\begin{bmatrix} \zeta_j^{yt} \\ \zeta_j^{xt} \\ \zeta_j^{wt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi^{yt} & & \\ 0 & \psi^{xt} & \\ 0 & \psi^{xwt} & \psi^{wt} \end{bmatrix} \right)$$

and

$$\begin{bmatrix} \zeta_j^{yc} \\ \zeta_j^{xc} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi^{yc} & \\ 0 & \psi^{xc} \end{bmatrix} \right) \tag{30}$$

$$\begin{bmatrix} \varepsilon_{ij}^{yt} \\ \varepsilon_{ij}^{xt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \theta_\varepsilon^{yt} & \\ 0 & \theta_\varepsilon^{xt} \end{bmatrix} \right). \tag{31}$$

In the control arm, the endogenous predictor's variance (ψ^{xc} in Equation (30)) needs to be estimated at (or above) the level where the predictor's effect is estimated; hence, b^{yc} appears at the cluster level in Equation (29). Importantly, because all predictors are endogenous in this specification, there is now no need to recode design-induced missingness on w in the manner described in Footnote 9.

For SSEM-PN, implementing the endogenous latent cluster mean approach for effect decomposition involves first obtaining latent within and between components of x_{ij}. To get the latent between component of x_{ij}, we treat the xs as endogenous indicators of their own latent variable (i.e., latent cluster mean of x, or random intercept of x, in the treatment arm). Specifically, elements of $\mathbf{x}_j^t$ (a $n_{t_1} \times 1$ vector here consisting of $x_{1j}^t - x_{5j}^t$) serve as indicators of a latent cluster mean in the treatment arm, denoted η_j^{xt}, and are coded missing-by-design in the control arm. x_{1j}^c serves as a single indicator of a corresponding latent variable, η_j^{xc}, in the control arm and is coded missing-by-design in the treatment arm. Then η_j^{xt} can be used as a predictor, in lieu of $\bar{x}_{\cdot j}$, to obtain the between effect of x_{ij}. To get the within component of x_{ij}, we employ a set of n_{t_1} latent variables, η_j^{xt}, wherein each latent variable has a single indicator, x_{ij}. Elements of η_j^{xt} are used as predictors to get a within effect of x_{ij}.

The full matrix expression for the SSEM-PN with endogenous latent cluster mean is given in Equation (32). This specification is diagrammed in Figure 5, Panel B. In Equation (32), all residuals are compactly represented in $\boldsymbol{\zeta}_j$. The matrices in Equation (32) have their elements defined in Equations

(33a) and (33b). We continue to suppose $n_{t_1} = 5$.

$$\mathbf{y}_j = \boldsymbol{\Lambda} \boldsymbol{\eta}_j$$
$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \mathbf{B} \boldsymbol{\eta}_j + \boldsymbol{\zeta}_j \tag{32}$$

$$\begin{bmatrix} y_{1j}^c \\ \mathbf{y}_j^t \\ x_{1j}^c \\ \mathbf{x}_j^t \\ w_j \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} \\ \mathbf{0}_{5\times1} & \mathbf{1}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{I}_{5\times5} & \mathbf{0}_{5\times5} \\ 0 & 0 & 1 & 0 & 0 & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} \\ \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{1}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times5} & \mathbf{I}_{5\times5} \\ 0 & 0 & 0 & 0 & 1 & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} \end{bmatrix} \begin{bmatrix} \eta_j^{yc} \\ \eta_j^{yt} \\ \eta_j^{xc} \\ \eta_j^{xt} \\ \eta_j^{wt} \\ \boldsymbol{\eta}_j^{yt} \\ \boldsymbol{\eta}_j^{xt} \end{bmatrix}$$
$$\tag{33a}$$

$$\begin{bmatrix} \eta_j^{yc} \\ \eta_j^{yt} \\ \eta_j^{xc} \\ \eta_j^{xt} \\ \eta_j^{wt} \\ \boldsymbol{\eta}_j^{yt} \\ \boldsymbol{\eta}_j^{xt} \end{bmatrix} = \begin{bmatrix} \mu^{yc} \\ \mu^{yt} \\ \mu^{xc} \\ \mu^{xt} \\ \mu^{wt} \\ \mathbf{0}_{5\times1} \\ \mathbf{0}_{5\times1} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & b^{yc} & 0 & 0 & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} \\ 0 & 0 & 0 & b_{2B}^{yt} & b_{1B}^{yt} & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} \\ 0 & 0 & 0 & 0 & 0 & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} \\ 0 & 0 & 0 & 0 & 0 & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} \\ 0 & 0 & 0 & 0 & 0 & \mathbf{0}_{1\times5} & \mathbf{0}_{1\times5} \\ \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times5} & b_{3W}^{yt}\mathbf{I}_{5\times5} \\ \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times5} & \mathbf{0}_{5\times5} \end{bmatrix} \begin{bmatrix} \eta_j^{yc} \\ \eta_j^{yt} \\ \eta_j^{xc} \\ \eta_j^{xt} \\ \eta_j^{wt} \\ \boldsymbol{\eta}_j^{yt} \\ \boldsymbol{\eta}_j^{xt} \end{bmatrix}$$

$$+ \begin{bmatrix} \zeta_j^{yc} \\ \zeta_j^{yt} \\ \zeta_j^{xc} \\ \zeta_j^{xt} \\ \zeta_j^{wt} \\ \boldsymbol{\varepsilon}_{j5\times1}^{yt} \\ \boldsymbol{\varepsilon}_{j5\times1}^{xt} \end{bmatrix}. \tag{33b}$$

$\mathbf{y}_j^t$ and $\mathbf{x}_j^t$ are $n_{t_1} \times 1$ vectors containing $y_{1j}^t - y_{5j}^t$ and $x_{1j}^t - x_{5j}^t$, respectively. η_j^{yc} and η_j^{xc} are latent components (factors) for y_{1j}^c and x_j^c. η_j^{xt}, η_j^{wt}, and η_j^{yt} are endogenous latent between components (factors) for $\mathbf{x}_j^t$, w_j, and $\mathbf{y}_j^t$, whereas $\boldsymbol{\eta}_j^{xt}$ and $\boldsymbol{\eta}_j^{yt}$ are endogenous latent within components (factors) for $\mathbf{x}_j^t$ and $\mathbf{y}_j^t$. In previous models, we did not explicitly define latent within components for the outcome in the treatment arm (i.e., $\boldsymbol{\eta}_j^{yt}$), but we do so here for consistency with how we are treating the predictors. $\mathbf{B}$ is a square matrix representing effects of endogenous latent variables on each other. In the treatment arm, the within effect of x_{ij} is b_{3W}^{yt} and its between effect is b_{2B}^{yt}. We can now directly estimate this between effect because we are including as a predictor the purely

within $\boldsymbol{\eta}_j^{xt}$ instead of the unsplit x_{ij} (see Curran et al., 2012). The between effect of w_j is $b_{1_B}^{yt}$. In the control arm, the effect of x_{ij} is b^{yc}. Deviations and residuals are distributed as

$$
\begin{bmatrix} \zeta_j^{yc} \\ \zeta_j^{yt} \\ \zeta_j^{xc} \\ \zeta_j^{xt} \\ \zeta_j^{wt} \\ \boldsymbol{\varepsilon}_{j_{5\times1}}^{yt} \\ \boldsymbol{\varepsilon}_{j_{5\times1}}^{xt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \mathbf{0}_{5\times1} \\ \mathbf{0}_{5\times1} \end{bmatrix}, \right.
$$

$$
\left. \begin{bmatrix} \psi^{yc} \\ 0 & \psi^{yt} \\ 0 & 0 & \psi^{xc} \\ 0 & 0 & 0 & \psi^{xt} \\ 0 & 0 & 0 & \psi^{xwt} & \psi^{wt} \\ \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \theta_\varepsilon^{yt}\mathbf{I}_{5\times5} \\ \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times1} & \mathbf{0}_{5\times5} & \theta_\varepsilon^{xt}\mathbf{I}_{5\times5} \end{bmatrix} \right) . \quad (34)
$$

In the control arm, we opted to estimate the predictor variance, ψ^{xc}, and outcome variance, ψ^{yc}, at the cluster level (as previously illustrated Figure 2, Panel B) for compactness of representation. Finally, using this SSEM-PN with now endogenous predictors, design-induced missingness on x and w in the control arm is handled by FIML and does not require the recoding described in Footnote 8.

SIMULATED EXAMPLE

In the previous section it was stated that, when using exogenous observed $\bar{x}_{.j}$, SEM approaches for handling partial nesting yield a specification of effect decomposition Strategies A and B in the treatment arm that is equivalent to, but arguably more straightforward than, MLM-PN. Here we illustrate this equivalency using a two-arm example data set with population parameters given in Table 3. These parameters are the same as in the two-arm Table 2 example with the exception that x_{ij} now has different within and between effects in the treatment arm. This sample had 90 treatment clusters of size 5, 350 control persons, a .30 ICC_y in the treatment arm, and homoscedastic residual variances across arms. x_{ij} and w_j were distributed as in the two-arm generating model from Table 2. Table 3 results show that, when decomposing the effect of x_{ij} in the treatment arm using $\bar{x}_{.j}$, we obtain equivalent results (parameter estimates, standard errors, and log-likelihood) between MLM-PN and MSEM-PN (for Strategy A) and between MLM-PN and SSEM-PN (for Strategy B).

The previous section also described how the SEM approaches both provide the option of substituting an endogenous latent cluster mean of x (i.e., random intercept) for $\bar{x}_{.j}$ in the treatment arm. Table 4 presents results of MSEM-PN and SSEM-PN with an endogenous latent cluster mean of x, fit to the same data set from Table 3. Comparing the Table 3 and Table 4 results illustrates that SEM approaches with a latent cluster mean of x can yield a more accurate estimate of the between effect of x_{ij} (and potentially other cluster-level parameters) because they are not subject to attenuation from use of the observed $\bar{x}_{.j}$.

Summary

We have now described two useful capabilities of SSEM-PN and MSEM-PN that are not accommodated by MLM-PN: cluster-level outcomes and decomposition of between versus within effects of person-level predictors in the treatment arm using endogenous latent cluster means (random intercepts). Both extensions had been discussed for SEM and MSEM in varying levels of detail in prior literature (e.g., Bauer, 2003; Curran, 2003; Curran et al., 2012; Lüdtke et al., 2008; McDonald, 1994; Preacher et al., 2010) but had not been adapted for the partially nested setting. We used two-paneled figures to juxtapose the SSEM-PN specification with observed $\bar{x}_{.j}$ versus latent cluster mean of x (top vs. bottom panels of Figure 5) and to juxtapose the MSEM-PN specification using observed $\bar{x}_{.j}$ versus latent cluster mean of x (top vs. bottom panels of Figure 4) when decomposing effects of person-level predictors in the treatment arm. Given the potential for different between versus within effects of x_{ij} in the treatment arm, researchers interested in testing the equality of the effect of x_{ij} *across arm* can consider a few options. For instance, researchers could test the equality of the simple effect in the control arm versus the total effect in the treatment arm, as done in the empirical example. In other substantive contexts researchers may be interested in testing the equality of the simple effect in the control arm versus the within effect (or between effect) in the treatment arm.

EMPIRICAL EXAMPLE

In this section, we illustrate the combination of our two extension topics (cluster-level outcomes and decomposition of person-level predictor effects) in SSEM-PN and MSEM-PN in the context of a model relevant to treatment-outcome studies. The empirical example data set is from an intervention study aimed at lowering the risk of depression and other internalizing and externalizing psychopathology among preadolescent and adolescent children (9–15 years) whose parent had a history of depression (see Compas et al., 2010; Compas et al., 2009; Compas et al., 2011, for details). Children and their parents were randomly assigned to either a treatment arm, where a cognitive-behavioral intervention was administered in therapy groups, or a control arm, where written information packets were sent by mail. The group therapy intervention involved eight weekly sessions followed by four monthly sessions; a variety of measures were recorded for both arms at baseline and after 2, 3, 6, and 12 months. The analysis data set contains a subsample of 54 children in the control arm and 14 therapy groups in the treatment arm, with 4 children per therapy group. One child per family was

TABLE 3
Partial Nesting Model Decomposing the Effect of an Exogenous Covariate Using Observed *x* Mean:
Illustration of Strategy A and B With MLM-PN, SSEM-PN, and MSEM-PN for a Simulated Single-Sample Data Set

Parameters Description	Pop-Gen Val.	Strategy A Notation MLM-PN Eqn 25	Strategy A Notation MSEM-PN Eqn 26	Results MLM-PN or MSEM-PN Est (SE)	Strategy B Notation MLM-PN Eqn 27	Strategy B Notation SSEM-PN Eqn 10b, 28, & Text	Results MLM-PN or SSEM-PN Est (SE)
Intercept, cont arm	4	γ_{00}	μ^c	4.04 (.09)	γ_{00}	μ^c	4.04 (.09)
Intercept, tx arm†	1	$\gamma_{00} + \gamma_{10}$	μ^t	1.06 (.12)	$\gamma_{00} + \gamma_{10}$	μ^t	1.06 (.12)
Condit. tx effect†	−3	γ_{10}	$\mu^t - \mu^c$	−2.98 (.15)	γ_{10}	$\mu^t - \mu^c$	−2.98 (.15)
L2 res. var, tx arm	.75	τ_{11}	ψ^t	.77 (.18)	τ_{11}	ψ^t	.77 (.18)
Res. var, cont arm*	2	σ_c^2	θ_ε^c	2.06 (.12)	σ_c^2	θ_ε^c	2.06 (.12)
L1 res. var, tx arm*	2	σ_t^2	θ_ε^t	2.06 (.12)	σ_t^2	θ_ε^t	2.06 (.12)
Effect x_{ij}, cont arm	1	γ_{30}	κ^c	.95 (.09)	γ_{30}	κ^c	.95 (.09)
Within-effect x_{ij}, tx arm	.75	γ_{20}	κ_W^t	.71 (.09)	γ_{20}	κ_W^t	.71 (.09)
Betwn-effect x_{ij}, tx arm†	.25	γ_{12}	ω_{2B}^t	.45 (.18)	$\gamma_{12} + \gamma_{20}$	$\omega_2^t + \kappa_W^t$	.45 (.18)
Contextual-effect x_{ij}, tx arm†	−.5	$\gamma_{12} - \gamma_{20}$	$\kappa_B^t - \kappa_W^t$	−.27 (.20)	γ_{12}	ω_2^t	−.27 (.20)
Effect w_j, tx arm	.50	γ_{11}	ω_1^t	.59 (.10)	γ_{11}	ω_1^t	.59 (.10)
Log-likelihood				−1294.23			−1294.23

Note. *Indicates held equal to obtain homoscedastic residual variances. Strategy A and B were defined in the text. Between and contextual effects of *x* control for *w*; the effect of *w* controls for the *x* mean. †Compound coefficients (sums or differences of estimated parameters) have *SE*s computed via the delta method using a model constraint command in Mplus or an estimate statement in SAS. See Table 1 note for abbreviation definitions.

included. Participant-initiated (not design-induced) missing data were handled via multiple imputation.

Analysis variables were parent perceptions of familial stress at 6 months postbaseline (Responses to Stress Questionnaire; Connor-Smith, Compas, Wadsworth, Thomsen, & Saltzman, 2000), child internalizing symptoms at 12 months postbaseline (Youth Self Report; Achenbach & Rescorla, 2001), and average amount of cognitive-behavioral homework[15] completed by a therapy group's members. Due to 6 months of therapy-group interaction, perceived familial stress reports were correlated within therapy group in the treatment arm (ICC$_x$ = .27). That is, some therapy groups on average report higher stress than others.

Our research questions pertain to three topics: testing moderation by treatment, decomposing effects in the treatment arm, and testing mediation at the cluster level in the treatment arm. Question (1): Is the total effect of parent perceptions of familial stress at 6 months (x_{ij}) on child internalizing symptoms at 12 months (y_{ij}) moderated by treatment? If treatment serves to decrease the impact of familial stress on symptoms, there may be a stronger effect of x_{ij} on y_{ij} in the control than treatment arm. Question (2): In the control arm, does a child's internalizing score change as a function of stress? Question (3a): In the treatment arm, does a child's internalizing score change as a function of stress, relative to his or her therapy-group mean (i.e., does the within effect = 0)? Question (3b): In the treatment arm, does the therapy-group average internalizing score change as a function of

group-average stress (i.e., does the between effect = 0)? Question (4): Is the between effect of stress on child internalizing symptoms in the treatment arm mediated by a cluster-level variable: the average amount of cognitive-behavioral homework completed by a therapy group's members (m_j)? We hypothesized that, first, therapy groups in which members were more stressed may complete fewer homework assignments. Second, we hypothesized that completing more homework assignments may reduce the risk of child internalizing symptoms at 12 months. Hence, homework completion was examined as a mediator of the link between family stress and child internalizing problems only in the treatment arm.

The empirical example model used to investigate these questions is depicted in path diagrams in Figures 6 and 7 in SSEM-PN and MSEM-PN versions. These versions are likelihood equivalent. Figure 6 depicts the SSEM-PN for one generic cluster per arm, and Figure 7 depicts the MSEM-PN for one generic person per arm. Symbol labels in Figures 6 and 7 correspond to the empirical example SSEM-PN and MSEM-PN equations in the online Appendix. To address research questions (3a) and (3b), the empirical example SSEM-PN and MSEM-PN decompose the person-level effect of x_{ij} into within and between effects using an endogenous latent cluster mean of *x* (i.e., random intercept of *x*). Taken together, this empirical example model involves a latent cluster mean of *x* and involves structural relations among multivariate outcomes (including a cluster-level outcome, m_j, measured only in the treatment arm) and so cannot be fit with MLM-PN. Our empirical example model happens to be saturated; otherwise we could evaluate absolute model fit.

---

[15]This predictor was divided by 10 to render its scale more comparable to scales of the other variables.

TABLE 4
Partial Nesting Model Decomposing the Effect of an Endogenous Covariate Using a Latent Cluster Mean:
Illustration of SSEM-PN and MSEM-PN for the Single-Sample Data Set From Table 3

Description	Pop-Gen. Values	Notation Eqn 29–31 & Eqn 32–34	MSEM-PN w/Latent Cluster Mean Results Est (SE)	SSEM-PN w/Latent Cluster Mean Results Est (SE)
Intercept, cont arm	4	μ^{yc}	4.04 (.09)	4.04 (.09)
Intercept, tx arm†	1	μ^{yt}	1.05 (.12)	1.05 (.12)
Condit. tx effect†	−3	$\mu^{yt} - \mu^{yc}$	−2.99 (.15)	−2.99 (.15)
L2 res. var, tx arm	.75	ψ^{t}	.75 (.18)	.75 (.18)
Res. var, cont arm*	2	ψ^{c}	2.06 (.12)	2.06 (.12)
L1 res. var, tx arm*	2	θ_{ε}^{t}	2.06 (.12)	2.06 (.12)
Effect x_{ij}, cont arm	1	b^{yc}	.95 (.09)	.95 (.09)
Within-effect x_{ij}, tx arm	.75	b_{3W}^{yt}	.71 (.10)	.71 (.10)
Betwn-effect x_{ij}, tx arm	.25	b_{2B}^{yt}	.32 (.26)	.32 (.27)
Effect w_{j}, tx arm	.50	b_{1B}^{yt}	.59 (.11)	.59 (.10)
Log-likelihood			−2371.39	−2371.39

Note. *Indicates held equal to obtain homoscedastic residual variances. In these fitted SEMs, both x and w are endogenous. The between effect of x is conditional on w (and vice versa). †Compound coefficients (sums or differences of estimated parameters) have *SE*s computed via the delta method using model constraint in Mplus or an estimate statement in SAS Proc Mixed. See Table 1 note for abbreviation definitions. Strategies A and B were defined in the text.

Results pertaining to our research questions are discussed in turn. (1) We rejected the null hypothesis that the unconditional total (conflated) effect of family stress on child internalizing symptoms in the treatment arm (obtained by constraining $(b_{2B}^{yt} + b_{1B}^{yt} b_{1B}^{mt}) = b_{3W}^{yt}$) was equal to the simple effect of family stress on child internalizing symptoms in the control arm (i.e., b^{yc}), $p < .01$. Hence, we have evidence that this effect is moderated by treatment. (2) Estimates from the fitted model in Table 5 show that, in the control arm, family stress at 6 months significantly predicted child internalizing symptoms at 12 months ($b^{yc} = .83$, $p < .05$). (3a)–(3b) In

the treatment arm, neither the unconditional within effect ($b_{3W}^{yt} = -.12$, $p > .05$) nor between effect ($b_{2B}^{yt} + b_{1B}^{yt} b_{1B}^{mt} = .06$, $p > .05$) of family stress on child internalizing symptoms significantly departed from 0. Although there are many possible explanations for this finding, one is that treatment may have suppressed the effects of family stress on child internalizing symptoms. (4) Finally, we investigated whether the amount of homework completed mediated the between effect of family stress on child internalizing symptoms in the treatment arm. As hypothesized, completing more homework was associated with a lower average child internalizing score
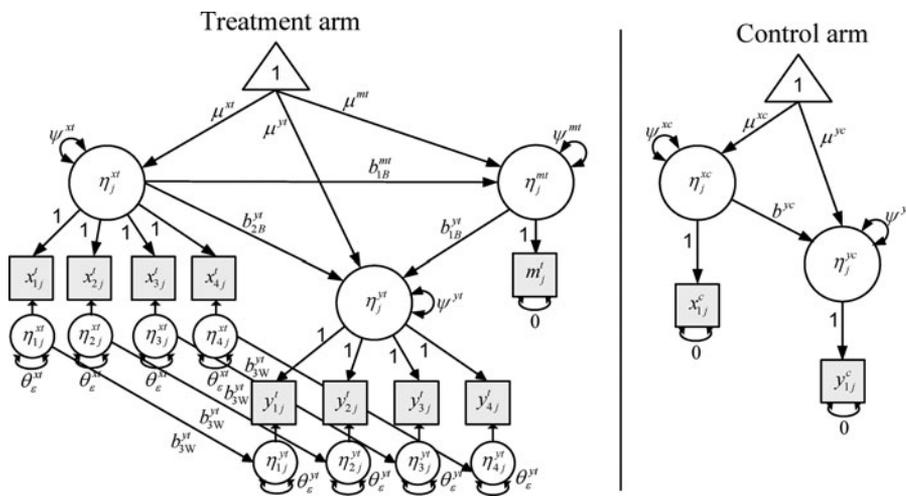

FIGURE 6 Depression intervention empirical example SSEM-PN: Incorporates a cluster-level outcome (m_j) and decomposes a person-level predictor's (x_{ij}) effect in the treatment arm using an endogenous latent cluster mean. *Note.* x_{ij} is parent perception of familial stress (at 6 months); y_{ij} is child internalizing (at 12 months); m_j is a therapy-group's average amount of completed homework. Diagram depicts the jth cluster in the treatment arm (left side) or control arm (right side). All predictors are endogenous. Notation in the figure corresponds to the empirical example SSEM-PN whose equations are given in the online Appendix. To simplify presentation, residuals ($\varepsilon_{ij}^{yt}, \varepsilon_{ij}^{xt}, \zeta_{j}^{yt}, \zeta_{j}^{yc}, \zeta_{j}^{mt}, \zeta_{j}^{xt}, \zeta_{j}^{xc}$) are not shown; only their variances are shown. SSEM-PN = multivariate single-level structural equation model for partial nesting.
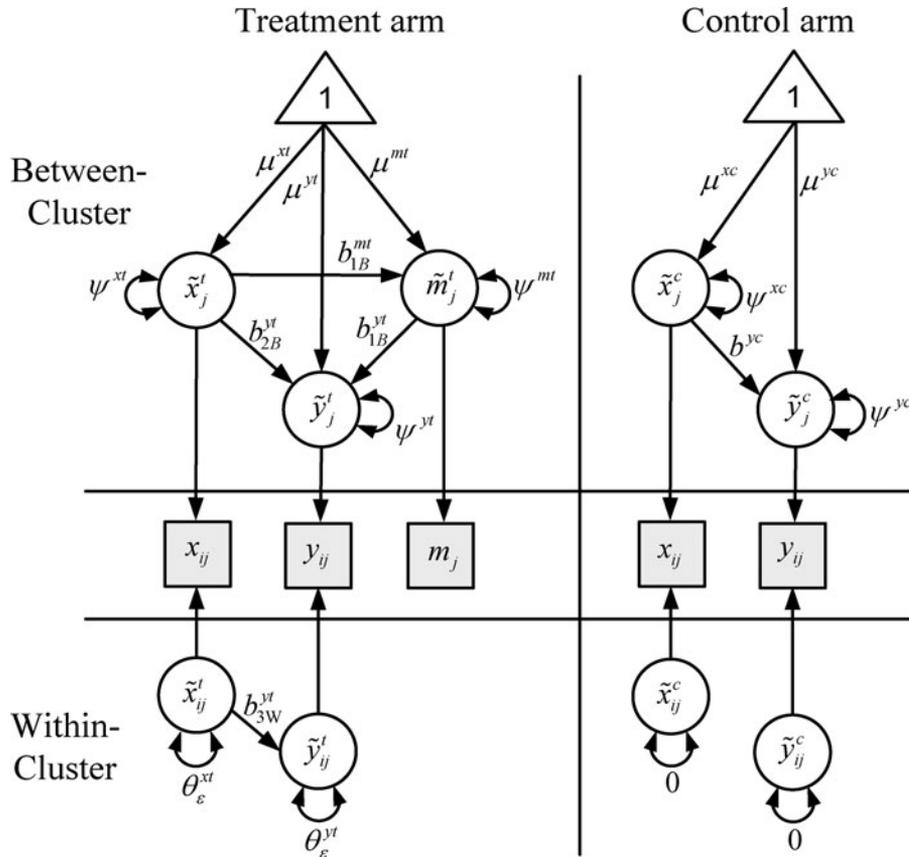
FIGURE 7 Depression intervention empirical example MSEM-PN: Incorporates a cluster-level outcome (m_j) and decomposes a person-level predictor's (x_{ij}) effect in the treatment arm using an endogenous latent cluster mean. *Note.* x_{ij} is parent perception of familial stress (at 6 months); y_{ij} is child internalizing (at 12 months); m_j is a therapy-group's average amount of completed homework. Diagram depicts the ith person in the jth cluster in the treatment arm (left side) or control arm (right side). Notation in the figure corresponds to the empirical example MSEM-PN whose equations are given in the online Appendix. To simplify presentation, residuals (ε_{ij}^{yt}, ε_{ij}^{xt}, ζ_j^{yt}, ζ_j^{yc}, ζ_j^{mt}, ζ_j^{xt}, ζ_j^{xc}) are not shown; only their variances are shown. MSEM-PN = multiple-arm multilevel structural equation model for partial nesting.

($b_{1B}^{yt} = -1.01, p < .05$). Contrary to our hypothesis, greater familial stress did not significantly affect the amount of homework completed by a therapy group ($b_{1B}^{mt} = -.07$, $p > .05$). Hence, amount of homework completed was not a significant mediator at the therapy-group level (a Monte Carlo[16] 95% confidence interval for the indirect effect, $b_{1B}^{yt} b_{1B}^{mt}$, included 0: {−.92, 1.17}). This finding may highlight the importance of enhancing homework completion at the therapy-group level; doing so could potentially improve child internalizing scores. Finally, as shown in Table 5, in the treatment arm, significant unexplained variability in child internalizing symptoms remains at the person level, and significant unexplained variability in average homework completed remains at the therapy-group level. In the control arm, unexplained variability in child internalizing symptoms also exists. Fu-

ture research can seek to explain these remaining sources of variance with additional predictors.

Summary

In this example, SSEM-PN and MSEM-PN allowed testing whether treatment moderates the effect of family stress on child internalizing symptoms and allowed examining cluster-level-specific mechanisms of change in the treatment arm. Empirical findings indicated that, in the control arm but not treatment arm, familial stress may increase child internalizing symptoms. Also, the more homework a therapy group completed, the better their average child internalizing scores.

DISCUSSION

Partial nesting designs are widely used in treatment outcome studies, particularly in clinical and educational settings (Sanders, 2011). Recent methodological advances for analyzing data from partially nested designs have used a multilevel modeling framework (MLM-PN; Bauer et al., 2008; Lee

---

[16]In multilevel (or partially nested) settings, confidence intervals for indirect effects can be obtained using the Monte Carlo method described in Preacher and Selig (2012) and MacKinnon, Lockwood, and Williams (2004).

TABLE 5
MSEM-PN and SSEM-PN Results for Depression
Intervention Empirical Example Analysis

Description	Notation† MSEM-PN or SSEM-PN	Est (SE)
Treatment arm		
Intercept y_{ij}, tx arm	μ^{yt}	13.21 (7.52)
Within effect of x_{ij} on y_{ij}, tx arm	b^{yt}_{3W}	−.12 (.22)
Between direct effect of x_{ij} on y_{ij}, tx arm‡	b^{yt}_{2B}	−.03 (.64)
Between effect of m_j on y_{ij}, tx arm‡	b^{yt}_{1B}	−1.01 (.47)*
Intercept m_j, tx arm	μ^{mt}	6.98 (3.91)
Between effect of x_{ij} on m_j, tx arm	b^{mt}_{1B}	−.07 (.38)
Level 1 res. var y_{ij}, tx arm	$\theta^{yt}_{\varepsilon}$	17.98 (6.36)*
Level 2 res. var y_{ij}, tx arm	ψ^{yt}	3.19 (4.63)
Level 2 res. var m_j, tx arm	ψ^{mt}	3.28 (1.31)*
Control arm		
Intercept y_{ij}, control arm	μ^{yc}	.04 (3.31)
Effect of x_{ij} on y_{ij}, control arm	b^{yc}	.83 (.28)*
Res. var y_{ij}, control arm	ψ^{yc}	67.99 (15.59)*
Log-likelihood		−712.25

Note. †Notation corresponds with Figures 6 and 7 and with online Appendix equations for the empirical example. * = $p < .05$. Empirical example SEMs used endogenous predictors and a latent cluster mean of x. ‡The between effect of x_{ij} on y_{ij} controls for m_j, and vice versa. Means and variances of x_{ij} are not usually of substantive interest and so are not tabled [mean and variance of x_{ij} in the control arm: 12.03(.80); 28.15(7.90); mean, Level 1 variance, and Level 2 variance of x_{ij} in the tx arm: 10.27(.89); 16.27 (4.37); 5.92 (5.55)]. See Table 1 note for abbreviation definitions for tx; res.; MSEM-PN; SSEM-PN. Condit. = conditional.

& Thompson, 2005; Moerbeek & Wong, 2008; Roberts & Roberts, 2005). Researchers analyzing data from treatment-outcome studies may nevertheless desire the flexibility of an SEM framework for including complex structural models, traditional factor predictors or outcomes, or cluster-level outcomes (e.g., Coatsworth, Duncan, Greenberg & Nix, 2010). In this article, a multivariate single-level SEM (SSEM-PN) and a multiple-arm multilevel SEM (MSEM-PN) for handling partially nested data were introduced. They were shown to reproduce results from existing MLM-PNs when MLM-PNs are applicable (with, e.g., homoscedastic vs. heteroscedastic residual variances across-arm, person-level and cluster-level covariates, and multiple treatment arms). Equivalencies of parameters across models were summarized in tables (see Tables 1 and 2) and were demonstrated with simulated data. We then described how absolute model fit, using a modified saturated model, is available for some SSEM-PNs and MSEM-PNs (which require only random intercepts) but are unavailable for all MLM-PNs (which require a random slope). This article provided the first thorough explanation of options for decomposing person-level predictor effects in the treatment arm of a partially nested design. Whereas

MLM-PN provides one option for doing so, MSEM-PN and SSEM-PN provide two options—using observed or latent cluster means. We illustrated both options with simulated data in Tables 3 and 4. Several SSEM-PN and MSEM-PN extensions were combined in the context of an empirical example analysis of a partially nested depression intervention data set. Software syntax for all fitted models was provided in the online Appendix.

Table 6 serves as a convenient summarization tool to compare and contrast how each of the three specifications, MLM-PN, SSEM-PN, and MSEM-PN, handles the unique features and modeling requirements of partially nested data. Table 6 also summarizes where data management is employed, particularly involving handling of control arm missingness-by-design, as a prelude to model fitting.

When choosing among MLM-PN, SSEM-PN, and MSEM-PN in practice, one strategy is to use the modeling framework with which one is more familiar, unless particular modeling features are desired that are better accommodated by another framework. Other considerations are as follows. Given the multivariate (wide-format) data structure of SSEM-PN and the long-format data structure of MSEM-PN and MLM-PN, the latter models are easier to specify when there are many persons per cluster in the treatment arm. In common partial nesting design settings, such as when treatment clusters are therapy groups, there are relatively few persons within cluster in the treatment arm (Baldwin et al., 2011), so this would not be a deciding factor. Also, it would be more straightforward to include random slopes in MSEM-PN or MLM-PN than in SSEM-PN (which would require use of definition variables; see Mehta & Neale, 2005). In current software, including more than two levels of (full or partial) nesting is easier in MSEM-PN than SSEM-PN, and more than three are easiest in MLM-PN. A relative advantage of SSEM-PN or MLM-PN is that they can be fit in a wider variety of software, whereas MSEM-PN requires a specialized software package that permits multilevel SEM.

Finally, we did not discuss at least one other alternative SEM specification for partial nesting (other than SSEM-PN, MSEM-PN) because it has a less straightforward implementation and less estimation stability. This alternative involves fitting a multivariate (wide-format) SEM with a fixed intercept and a random slope for treatment. The first loading on the random slope factor could be a definition variable (allowed to be 1 for treated persons, 0 for control persons), and remaining indicators (as many as the maximum number of persons per cluster) could be 1 for treated persons and missing otherwise. Among other limitations, this approach requires an inconvenient amount of programming to allow heteroscedastic residual variances across arms (i.e., it requires implementation of developments in B. O. Muthén & Asparouhov [2002] with the identification constraint modified as in Snijders & Bosker [1999, sect. 8.1] or Goldstein, [2003, sect. 3.1.2] to enable the outcome residual variance to be either larger or smaller in the treatment vs. control arms).

TABLE 6
Summary of How MLM-PN, SSEM-PN, and MSEM-PN Each Handle Modeling Requirements of Partially Nested Data

Unique Feature of Partially Nested Data	How Feature Is Handled by Alternative Specifications		
	MLM-PN (Long Format)	SSEM-PN (Wide Format; $n_{t_1} = 5$)	MSEM-PN (Long Format)
Outcome variance split into between/within components in treatment arm only. May need heterogeneous residual variances across arm.	Treatment toggle variable ($0 = $ control, $1 = $ treated) multiplied by random effect desired to only pertain to treatment arm. Residual variance can depend on arm.	Modify wide-format data: split y_{1j} into y^c_{1j} and y^t_{1j}, y^c_{1j} missing for treated; y^t_{1j}–y^t_{5j} missing for controls. Include random intercept of y with indicators y^t_{1j}–y^t_{5j}; estimate its factor and residual variances. Estimate variance for y^c_{1j} at one level.	Multiple-group specification (where group $=$ arm) allows decomposition of y variance into between/within components in treatment but not control arm.
L2 outcome may be in treatment arm only.	Not applicable.	Include L2 outcome (missing for controls).	Include L2 outcome in treatment arm. Constrain its parameters to 0 in control arm.
L2 covariate may only require effect in treatment arm.	Product term of L2 covariate and treatment toggle variable included. No main effect of L2 covariate included.	L2 covariate predicts random intercept of y.	L2 covariate predicts between component of y in treatment arm.
L2 covariate may be missing-by-design in control arm.	For controls, assign arbitrary numeric code (not missing data code) for exogenous L2 covariate missing-by-design.	**Exogenous covariate:** Same as MLM-PN for L2 (& L1) covariates missing-by-design.	**Exogenous covariate:** Same as MLM-PN.
		Endogenous covariate: Nothing.	**Endogenous covariate:** Nothing.
Decompose effect of L1 covariate (x_{ij}) in treatment arm but not control arm.	Include as predictors, for Strategy A: x_{ij}; $\bar{x}_{.j}$; product of x_{ij} with a control toggle variable. For Strategy B: $\bar{x}_{.j}$; product of x_{ij} with each treatment and control toggle variables.	Strategy B. $\bar{x}_{.j}$ predicts random intercept of y. Within effect of x_{ij} on y^t_{1j} need not be equated to x_{ij} on y^c_{1j}.	Strategy A. Include as predictors: x_{ij} and $\bar{x}_{.j}$ in treatment arm and x_{ij} in control arm.
		Split x_{ij} into x^c_{1j} and x^t_{1j}–x^t_{5j} missing for treated; x^t_{1j}–x^t_{5j} missing for controls. $\bar{x}_{.j}$ replaced by random intercept of x with indicators x^t_{1j}–x^t_{5j}. Within effect of x^t_{ij} on y^t_{ij} need not be equated to x^c_{1j} on y^c_{1j}.	Decompose effect of x_{ij} into latent within and between components in treatment arm but only a single component in control arm.

Note. This table uses 2 arms. Exogenous covariates require a conditional likelihood. L2 $=$ Level 2; L1 $=$ Level 1. Abbreviations defined in Table 1 note.

Other potential extensions of SSEM-PN or MSEM-PN for partial nesting designs that were not discussed here present interesting avenues for future work. For instance, in the empirical example, m_j, the average amount of homework completed by a therapy group's members, could instead have been replaced by a traditional factor with multiple indicators at the cluster level. For instance, it could be replaced by a latent motivation factor with indicators of: therapy group average homework completed, perceived investment or belief in treatment, and amount of therapist encouragement or rewards for completing homework. For such a model, absolute model fit would be useful to calculate (as described earlier) and interpret. Second, measured variable outcomes that are categorical rather than continuous could be considered (e.g., depression diagnosis rather than symptom subscale score). Third, categorical (e.g., binary) x_{ij} or w_j could be employed, although estimation would be less computationally intensive if they are considered exogenous rather than endogenous. Fourth, SSEM-PN or MSEM-PN could be used in sample size planning for optimal power, as has been done with MLM-PN (Candel & van Breukelen, 2009, 2010; Moerbeek & Wong, 2008).

In conclusion, the SSEM-PN and MSEM-PN specifications, interpretation, and simulated and empirical demonstrations provided here allow researchers greater flexibility for analyzing data from partial nesting designs in an SEM framework. We encourage future work extending SSEM-PN and MSEM-PN to include more complex SEM specifications relevant to treatment-outcome research.

FUNDING

REFERENCES

Achenbach, T., & Rescorla, L. (2001). *Manual for ASEBA School-Age Forms and Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Asparouhov, T., & Muthén, B. (2007). *Constructing covariates in multilevel regression (*Web *Note 11)*. Retrieved from http://www.statmodel.com

Baldwin, S., Bauer, D., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. *Psychological Methods*, *16*, 149–165.

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.

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

Becker, S., Mlay, R., & Schwandt, H. (2010). Comparing couples' and individual voluntary counseling and testing for HIV at antenatal clinics in Tanzania: A randomized trial. *AIDS Behavior*, *14*, 558–566.

Beran, T., & Violato, C. (2010). Structural equation modeling in medical research: A primer. *BMC Research Notes*, *3*, 1–10.

Candel, M., & van Breukelen, G. (2009). Varying cluster sizes in trials with clusters in one treatment arm: Sample size adjustments when testing treatment effects with linear mixed models. *Statistics in Medicine*, *28*, 2307–2324.

Candel, M., & van Breukelen, G. (2010). *D*-optimality of unequal versus equal cluster sizes for mixed effects linear regression analysis of randomized trials with clusters in one treatment arm. *Computational Statistics and Data Analysis*, *54*, 1906–1920.

Coatsworth, J., Duncan, L., Greenberg, M., & Nix, R. (2010). Changing parent's mindfulness, child management skills and relationship quality with their youth: Results from a randomized pilot intervention trial. *Journal of Child and Family Studies*, *19*, 203–217.

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin*, *114*, 174–184.

Compas, B. E., Champion, J. E., Forehand, R., Cole, D. A., Reeslund, K. L., Fear, J., . . . Roberts, L. (2010). Coping and parenting: mediators of 12-month outcomes of a family group cognitive-behavioral preventive intervention with families of depressed parents. *Journal of Consulting and Clinical Psychology*, *78*, 623–634.

Compas, B. E., Forehand, R., Keller, G., Champion, A., Reeslund, K. L., McKee, L., . . . Cole, D. A. (2009). Randomized clinical trial of a family cognitive-behavioral preventive intervention for children of depressed parents. *Journal of Consulting and Clinical Psychology*, *77*, 1009–1020.

Compas, B. E., Forehand, R., Thigpen, J. C., Keller, G., Hardcastle, E. J., Cole, D. A., . . . Roberts, L. (2011). Family group cognitive-behavioral preventive intervention for families of depressed parents: 18- and 24-month outcomes. *Journal of Consulting and Clinical Psychology*, *79*, 488–499.

Connor-Smith, J., Compas, B., Wadsworth, M., Thomsen, A., & Saltzman, H. (2000). Responses to stress in adolescence: Measurement of coping and involuntary responses to stress. *Journal of Consulting and Clinical Psychology*, *68*, 976–992.

Croon, M. A., & van Veldhoven, J. P. (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.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619.

Curran, P. J., Lee, T., Howard, A., Lane, S., & MacCallum, R. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In J. Harring & G. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 217–253). Charlotte, NC: Information Age.

Dannon, P., Gon-Usishkin, M., Gelbert, A., Lowengrub, K., & Grunhaus, L. (2004). Cognitive behavioral group therapy in panic disorder patients: The efficacy of CBGT versus drug treatment. *Annals of Clinical Psychiatry*, *16*, 41–46.

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

Goldstein, H. (1995). *Multilevel statistical models (*2nd *ed.)*. New York, NY: Wiley.

Goldstein, H. (2003). *Multilevel statistical models (*3rd *ed.)*. New York, NY: Oxford.

Haugli, L., Steen, E., Laerum, E., Nygard, R., & Finset, A. (2001). Learning to have less pain: Is it possible? A one-year follow-up study of the effects of a personal construct group learning program on patients with chronic musculoskeletal pain. *Patient Education and Counseling*, *45*, 111–118.

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

Heller-Boersma, J., Schmidt, U., & Edmonds, D. (2007). A randomized controlled trial of a cognitive-behavioural group intervention versus waiting-list control for women with uterovaginal agenesis (Mayer–Rokitansky–Kuster–Hauser syndrome: MRKH). *Human Reproduction*, 22, 2296–2301.

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

Hoyle, R., & Smith, G. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology*, 62, 429–440.

Jarrett, R., Schaffer, M., McIntire, D., Witt-Browder, A., Kraft, D., & Risser, R. (1999) Treatment of atypical depression with cognitive therapy or phenelzine: A double-blind, placebo-controlled trial. *Archives of General Psychiatry*, 56, 431–437.

Jöreskog, K. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.

Jöreskog, K., & Sörbom, D. (2006). *LISREL 8.8 for Windows [Computer software]*. Lincolnwood, IL: Scientific Software International.

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions (2*nd ed.). Newbury Park, CA: Sage.

Kim, S.-Y., Mun, E.-Y., & Smith, S. (in press). Using mixture models with known class membership to address incomplete covariance structures in multiple-group growth models. *British Journal of Mathematical and Statistical Psychology*. doi:10.1111/bmsp.12008

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

Korendijk, E., Maas, C., Hox, J., & Moerbeek, M. (2012). The robustness of the parameter and standard error estimates in trials with partially nested data: A simulation study. In E. Korendijk, *Robustness and optimal design issues for cluster randomized trials* (pp. 59–94). Dissertation, Utrecht University.

Kreft, I. 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.

Ladouceur, R., Dugas, M., Freeston, M., Leger, E., Gagnon, F., & Thibodeau, N. (2000). Efficacy of a cognitive-behavioral treatment for generalized anxiety disorder: Evaluation in a controlled trial. *Journal of Consulting and Clinical Psychology*, 68, 957–964.

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

Leyland, A., & Goldstein, H. (2001). *Multilevel modeling of health statistics*. New York, NY: Wiley.

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

Lüdtke, O., Marsh, H., 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.

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.

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

McDonald, R., & Goldstein, H. (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.

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

Muthén, B. O. (1987). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model. Theoretical integration and user's guide*. Mooresville, IN: SSI.

Muthén, B. O. (1990, May). *Mean and covariance structure analysis of hierarchical data*. Presentation given at the annual meeting of the Psychometric Society, Princeton, NJ.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29, 81–117.

Muthén, B. O., & Asparouhov, T. (2002). *Modeling of heteroscedastic measurement errors (*Mplus *Web Note 3)*. Retrieved from www.statmodel.com

Muthén, B. O., & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In F. G. Davidian, M. G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC Press.

Muthén, L. K., & Muthén, B. O. (1998–2013). *Mplus users' guide. (*7th *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.

Pisinger, C., Vestbø, J., Borch-Johnsen, K., & Jørgensen, T. (2005). Smoking cessation intervention in a large randomised population-based study: The Inter99 study. *Preventive Medicine*, 40, 285–292.

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

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

Price, M., & Anderson, P. (2011). The impact of cognitive behavioral therapy on post event processing among those with social anxiety disorder. *Behaviour Research and Therapy*, 49, 132–137.

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

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

Russell, D., Kahn, J., Spoth, R., & Altmaier, E. (1998). Analyzing data from experimental studies: A latent variable structural equation modeling approach. *Journal of Consulting Psychology*, 45, 18–29.

Sanders, E. (2011). *Multilevel analysis methods for partially nested cluster randomized trials (*Unpublished *doctoral dissertation)*. Seattle: University of Washington.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Thomas, P., Thomas, S., Kersten, P., Jones, R., Nock, A., Slingsby, V., . . . Hillier, C. (2010). Multi-centre parallel arm randomised controlled trial to assess the effectiveness and cost-effectiveness of a group-based cognitive behavioral approach to managing fatigue in people with multiple sclerosis. *BMC Neurology*, 10, 1–12.

Thompson, L., Gallagher, D., & Breckenridge, J. (1987). Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology*, 55, 385–390.

Tomarken, A., & Waller, N. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*, 1, 31–65.

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

Van Minnen, A., Hoogduin, K., Keijsers, G., Hellenbrand, I., & Hendriks, G. (2003). Treatment of trichotillomania with behavioral therapy or fluoxetine. *Archives of General Psychiatry*, 60, 517–522.

Widaman, K., Grimm, K., Early, D., Robins, R., & Conger, R. (2013). Investigating factorial invariance of latent variables across populations when manifest variables are missing completely. *Structural Equation Modeling*, 20, 384–408.