

Using *n*-level Structural Equation Models for Causal Modeling in Fully Nested, Partially Nested,
and Cross-Classified Randomized Controlled Trials

Yaacov Petscher

Christopher Schatschneider

Florida State University

Published 2019

Petscher, Y., & Schatschneider, C. (2019). Using *n*-level structural equation models for causal modeling in fully nested, partially nested, and cross-classified randomized controlled trials.

Educational and Psychological Measurement, 79, 1075-1102

This research was supported by Grant R324A130262 from the Institute of Education Sciences. The content is solely the responsibility of the authors and does not necessarily represent the view of the Institute of Education Sciences.

Abstract

Complex data structures are ubiquitous in psychological research, especially in educational settings. In the context of randomized controlled trials, students are nested in classrooms but may be cross-classified by other units, such as small groups. Further, in many cases only some students may be nested within a unit while other students may not. Such instances of partial nesting requires a more flexible framework for estimating treatment effects so that the model coefficients are correctly estimate. Although several recommendations have been offered to the field on handling partially nested data, few are comprehensive in their treatment of manifest and latent variables in the context of partial nesting, full nesting, and cross-classification. The present study introduces *n*-level SEM (Mehta, 2013a) as a flexible measurement and analytic framework for the estimation of treatment effects for complex data structures that frequently present in randomized controlled trials. In this tutorial, we explore how the notation of *n*-level SEM allows for parsimonious model specification whether data are observed or latent and in the presence of partial nested or cross-classified designs. By using the *xxm* package in R, the advantage of using *n*-level SEM framework is demonstrated through five examples for single outcome manifest variables, as in the traditional multilevel model, as well as latent applications as in multilevel SEM.

Keywords: *n*-level SEM, multilevel SEM, partially nested data, cross-classified, randomized controlled trial

Using *n*-level Structural Equation Models for Causal Modeling in Fully Nested, Partially Nested,
and Cross-Classified Randomized Controlled Trials

Nested factors refer to the relation between two factors whereby the levels of one factor occur only under one level of another factor. One such example exists in the educational literature with students and schools factors, where each student is only present at one school. The students attending one school can be said to be nested or clustered at that school. Nested research designs are common in the social sciences (Baldwin, Bauer, Stice, & Rohde, 2011). These designs pose problems for many routine statistical models because the data from these designs often exhibit a lack of independence of observation, which is an assumption for most statistical models. Violating this assumption has serious impacts on the Type 1 error rate and the standard errors obtained from these models. To overcome such limitations, multilevel models were developed in part to address the needs to analyze data from nested designs. Multilevel (or mixed) models allow for a partitioning of the total variance into variance attributable either between or within clusters and allows the modeling of data from a number of different nested designs. Mixed models estimate the clustering effects and correct the estimates of the variances and standard errors of the model parameters.

Although mixed models have distinct advantages over other models when it comes to estimating effects from nested designs, one drawback is that they cannot estimate latent variables. Significant progress on this shortcoming has been made with the introduction of multilevel structural equation modeling (ML-SEM; Mehta & Neale, 2005). ML-SEM allows for “...full-blown SEM models to be developed at each level of nesting for clustered data” (Mehta & Neale, 2005, p. 259). The incorporation of ML-SEM into programs such as MPLUS has increased the usage of these models for a variety of designs – from individual differences studies

to cluster randomized control designs. Employing ML-SEM minimizes the influence of measurement error and allows the researcher focus on constructs of interest instead of measures of interest which could also reduce the Type I error rate by estimating relations among fewer constructs instead of observed measures. The reduction in measurement error will also increase the power of the model to detect effects.

Even with the advances in ML-SEM, shortcomings remain. Although they can fit a range of data with nested structures, their application is often limited in the number of hierarchical levels that can be fit, and mixed models are still more flexible in fitting complex data structures that have a mix of nested and crossed random factors. But one particular data structure that both mixed models and ML-SEM models have difficulty is designs that contain partial nesting. Partial nesting occurs when clustering happens under one or more levels of another factor, but not consistently across all levels. As an example, partial nesting would be observed in a treatment study if the participants in the treatment arm received intervention in small groups while the control arm did not. In this case, the small groups clustering factor only occurs under one level of the treatment factor. These types of designs are fairly common in the social sciences. Bauer, Sterba, and Halfors (2008) reviewed articles from four public health journals and found that 32% of the randomized control trials had a partially nested design; moreover, the partially nested design was more prevalent than the often-used fully nested design (27%). In a similar way, Sanders (2011) reviewed articles from four education journals and discovered that 15% of the randomized trials used a partially nested design.

Across these studies using partially nested designs, two analytic approaches have predominantly been used: 1) ignore the nesting altogether and utilize single-level regression, or 2) treat the data as a fully nested design with either a) individuals in the control arm set as being

nested in one broader group or b) artificially creating small groups in the control arm so that the design appears as fully nested. Neither of these approaches are acceptable. A single-level regression approach to partially nested data violates the independence assumption with Type I error rates differentially impacted according to the directionality of the within-group correlations (i.e., higher Type I error rates when Y is positively correlated among individuals within the treatment group and lower Type I error rates when Y is negatively correlated among individuals within a group; Bauer et al., 2008). Applying a fully nested design approach to partially nested data is also problematic and leads to biased estimates of treatment effects (Roberts & Roberts, 2005) due to inappropriate assumptions made about the decomposition of variance in the control group (i.e., $V(Y|Treatment = 0) = \tau_{00} + \sigma^2$) compared to the treatment group (i.e., $V(Y|Treatment = 1) = \tau_{00} + \sigma^2$) because the assumed grouping structures from a fully nested design (i.e., τ_{00}) does not apply to the control group in the partially nested design.

The response to the lack of alignment between experimental research design and statistical analysis of data has been a series of studies that have begun to unpack how statistical tools may be adapted to estimate unbiased treatment effects based on correct decomposition of individual- and cluster-level variances. Baldwin, Bauer, Stice, & Rohde, (2011) proposed some adaptations to fully nested models with manifest variables to accommodate partially nested designs. The authors demonstrated through simulation that using a Satterthwaite or Kenward-Rogers degrees of freedom maintains appropriate levels of Type I error rates with unbiased estimates of effects. Sterba et al (2014) presented two methodologies using structural equation models. The multivariate single-level structural equation model (SSEM) requires data manipulations of the dependent variable such that when data are in a wide format, y must be reorganized into $1 + n_t$ columns where n_t is the number of treatment arms. For example, with

one treatment condition, y is reorganized into two columns, one for the control condition (y_c) and one for the treatment condition (y_t) where missing data exists for individuals in the treatment condition on y_c and missing data exists for individuals in the control condition on y_t . The multiple-arm multilevel structural equation model (MSEM) approach uses data in a long file format with a multiple group approach to testing for differences between the treatment and control conditions.

The studies by Sterba et al. (2014) and Baldwin et al. (2011) have served the field by providing strategic insight into how statistical models may be used to address the partial nested condition, yet there is much more work to be done. By and large, most of the recommendations in the literature are focused on the mathematically underpinnings that show how data from partially nested designs in randomized controlled trials may be analyzed (e.g., Sterba, 2017). Further, even where these approaches are useful starters for supporting appropriate analytics matched to the design, there are a number of shortcomings in the approaches. First, many software packages do not have options for adjusting the estimator of the degrees of freedom as in the recommended approach from Baldwin et al. (2011). Second, the latent variable approaches outlined by Sterba et al. (2014) do not treat the data in its naturalistic format whereby additional data management via the creation of new variables and shifting data from wide-to-long are necessary means to achieve the analytic end. Third, no single paper can explicate or consider each and every variant of random effects modeling for partially nested data. One facet of these designs that has not yet been explored is the case of cross-classification that may occur with partial nesting. Although research has demonstrated the impacts of ignoring cross-classification with fully nested data via biased coefficients (Gilbert, Petscher, Compton, & Schatschneider, 2016; Luo & Kwock, 2009; Meyers & Beretvas, 2006), only one study has examined cross-

classification in the context of partially nested designs (Luo, Cappaert, & Ning, 2015). In the same manner that ignoring cross-classification in fully nested designs has shown to lead to biased coefficients, so too does cross-classification have bearing on partially nested designs. For example, within educational research where partial nested designs may use small groups for the treatment condition that involves student pull-outs from classrooms, the data structure is partially nested with cross-classification such that only treatment students are nested in small groups, but treatment and control students are nested in varying classrooms. It is important, therefore, to explore not only how data from partially nested designs may be more robustly analyzed but also how complex structures such as cross-classification should be accounted for as well.

What these three exemplar limitations highlight is that in some ways our mechanisms for modeling partially nested data are more restrained by software and less by the statistical and measurement underpinnings in the partial nested design. Until recently, with the introduction of *n*-level structural equation model (*n*SEM; Mehta, 2013a) there was no naturalistic multilevel SEM model solution to the problem of partially nested data that avoids differential estimation of degrees of freedom, data management, or creation of new variables. *n*SEM expands upon ML-SEM in its conceptualization of *levels*. In ML-SEM, level typically refers to a random factor that exists in a hierarchical relation with other factors, such as students nested in classrooms. *n*SEM broadened the definition of level to also refer to levels of a fixed effects factor. For example, sex may constitute two levels, one for males and one for females or in a randomized controlled trial with one treatment group and one control group each of groups would be a separate level in the model. In this framework, a much broader array of models can be estimated that can naturally handle complex data structures such as data from partially nested designs. The *n*SEM framework can a broad array of SEM models at each level of nesting or each level of a fixed

effect. In this way, designs with partially nested data can be directly modeled whereby the nested structure exists for one level of a fixed effect (treatment) but is absent from the other level of the factor (control). By treating the concept of level more flexibly in *n*SEM, it is possible that given an equivalence in results between an *n*SEM approach and approaches based on recommendations from prior studies (e.g., data manipulations and degrees of freedom estimator changes) an *n*SEM framework may render previous recommendations as unnecessary. With the attention that partial nesting has received in the literature (Baldwin et al., 2011; Bauer et al., 2008; Moerbeek & Wong, 2008; Sterba et al., 2013), our goal in this tutorial is to demonstrate the usefulness of the *n*SEM framework using the *xxM* package in R (Mehta, 2013b) across of number of different potential datasets where simple or complex clustering exists (e.g., cross-classification), especially where tests of causal inference are of importance.

Our tutorial is broadly concerned with the question: to what extent does an *n*SEM framework yield identical or similar results to conventional applications of partial nesting for a randomized controlled trial? We consider four manifest variable examples using a common data set: Example 1 compares a fully nested unconditional model of students nested within classrooms using *n*SEM and a traditional multilevel model to compare estimated means, variances, and confidence intervals; Example 2 builds on Example 1 by considering a fully nested version of a randomized controlled trial with one treatment group and one control group where students are nested in classrooms; Example 3 considers the partial nested randomized controlled group with the treatment students nested within small groups and the control group is not nested; Example 4 is the partially nested, cross-classified causal model where students are cross-classified according to small groups and classrooms and the treatment students only are

nested in small groups. Our tutorial concludes with Example 5 that teaches how ML-SEM with latent variables can be done with partial nesting and cross-classification.

To provide direct connections among the multilevel model, structural equation model, and *n*SEM framework, we begin with a brief refresher of the multilevel models and ML-SEM approaches to nested data and then we move directly to a demonstration of *n*SEM for fully nested, partially nested, and cross-classified models. Along the way, data and code are provided that the reader may use to replicate and extend to their own work.

Manifest Variable Modeling Frameworks

A well-known framework for modeling data from cluster randomized trials is multilevel modeling (MLM). For a given two-level model, such as students nested within classrooms, the unconditional model is typically represented by

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad \text{Eq. 1}$$

where Y_{ij} is a post-test score Y for student i in classroom j , γ_{00} is the grand mean intercept, u_{0j} represents the classroom deviations from the mean distributed as $N[0, \tau_{00}]$, and e_{ij} is the student deviation distributed as $N[0, \sigma^2]$. An alternative perspective for modeling data with the same nesting structure is structural equation models (SEM) and would represent the unconditional model as

$$Y_{ij} = \nu + (\lambda\eta_j) + e_{ij} \quad \text{Eq. 2}$$

where Y_{ij} and e_{ij} maintain identical interpretations to those found in Equation 1, ν represents the measurement intercept, $\lambda\eta_j$ is the linking of each Y_{ij} where λ is a fixed factor loading at 1.0, and η_j are the classroom means that are distributed $N[\alpha, \psi]$, where α is the grand mean intercept, and ψ is the variance (Mehta & Neale, 2005).

The equivalence between parameters in Eqs. 1 and 2 is provided in Figure 1 as a measurement cross-walk for the two general approaches. A careful look at Figure 1 shows many commonalities between the multilevel and SEM approaches to data clustering: shapes, lines, Greek letters, and subscript notations characterize both frameworks. Rectangles or squares represent the measured variable (Y_{ij}); circles can represent both latent factors that are the unmeasured construct of interest associated with the measured variable (β_{0j} and η_j), and errors associated with the measured variable (e_{ij}); triangles represent the mean structure for latent variables. Straight lines in the model represent paths linking predictor variables to predicted variables, such as the latent variable predicting the measured variable. Curved lines with arrows at both ends can represent disturbances when the line is associated with one shape (i.e., η_j , e_{ij} , u_{0j}) as well as covariances between either latent or measured variables as well as disturbances. Greek letters symbolize the various parameters in the model, such as those shown in Figure 1, and subscript notation is used to distinguish among different levels of clustering within one's data. Where the base mechanisms of the MLM and SEM diagrams are similar, differences are also seen. Grand mean intercepts, classroom deviations, and residual variances are represented with different parameters and subscript notations between the two representations. Classroom deviations are represented with different figures. Despite the equivalence of MLM and SEM when nesting exists (Curran, 2003), the differences in model representation and other noted limitations of each method (e.g., robustness of violations to assumptions, power, application of model constraints) may suggest that an alternative framework may be brought to bear generally for nested data and specifically for causal modeling.

Considering n -level SEM and xxm

An alternative framework to the general MLM and SEM approaches for nested data is that of *n*-level SEM (*n*SEM; Mehta, 2013a). *n*SEM represents a general structure that allows one to formulate and estimate complex data relations from cross-sectional and longitudinal studies rather simply. The burden of model specification in the presence of multiple levels of nesting, cross-classification, or multiple membership, coupled with complex nesting structures such as partial nesting, and multivariate outcomes, can be eased with an *n*SEM approach to model building which is, *if you can draw it, you can model it*. When estimated in software such as xxm (Mehta, 2013b), specified models have an exact equivalence among the graphical representation, the equations, the matrices, and the code. Consider the two-level example just presented in Figure 1 but expressed in Figure 2 through just the SEM and *n*SEM configurations. A careful comparison between the two shows identical structures with only parameter notation varying. Where the MLM and SEM equations and graphical representations include subscript notation to note clustering level they do not explicitly account for the number of variables in the model and the level at which the variable may belong. Rather than using the notation of *i* and *j* to denote unit clustering such as students and classrooms, *n*SEM utilizes subscript and superscript notation to denote variable number and level, respectively. Y_1^1 is the first variable (Y_1) at the first level (i.e., student level; Y^1), η_1^2 is the first latent variable at the second level (i.e., classroom) and are distributed as $N[\alpha_1^2, \psi_{1,1}^{2,2}]$ where α_1^2 is the grand mean for the first latent variable at the second level. $\psi_{p,p}^{u,u}$ is a variance of the *p*th latent variable at level *u*; thus, $\psi_{1,1}^{2,2}$ is the variance associated with η_1^2 . $\lambda_{v,q}^{u,p}$ is a factor loading from variable *q* at level *p* to variable *v* at level *u*; thus, $\lambda_{1,1}^{1,2}$ is the effect from η_1^2 to Y_1^1 and is fixed at 1.0. ν_1^1 is the measurement intercept for first variable at level 1 (Y_1^1) and is fixed at 1.0, and e_1^1 is the level-1 deviation distributed as $N[0, \theta_{1,1}^{1,1}]$, where $\theta_{p,p}^{u,u}$ is a

residual variance of the p th variable at level u . The general structure of the φ and θ parameters are such that they may be used expressed variances and covariances (φ) or residual variances and covariances (θ) (Mehta, 2013a).

The nSEM model parameters from Figure 2b may then be combined to form the equation of the exemplar two-level model as

$$Y_1^1 = \nu_1^1 + (\lambda_{1,1}^{1,2} \eta_1^2) + e_1^1. \quad \text{Eq. 3}$$

Note that outside of the superscript and subscript notations, Eqs. 2 and 3 are identical just as the measurement models for these two equations (Figures 2a and 2b) are nearly identical in structure. By broadening the traditional subscript notation to denote nesting in multilevel/SEM to one that uses subscripts and superscripts, nSEM moves from a conventional interpretation of complex structures as one of *nesting* to that of *influence* (Mehta & Petscher, 2016). Equations 1 and 2 primarily convey the relation between student and classrooms as one of structure where the former is nested within the latter, but Eq. 3 may be thought of as describing sources of influence from one level to another. This interpretation differences are largely semantic, but the implication of the changing from nesting to influence will becomes more apparent as the model complexity increases.

Although it may seem unnecessarily complex to construct equations with this extended notation, it will be seen that its intricacies leads to parsimonious measurement, more intuitive model construction process, and greater symmetry among equations, measurement models, and code that preserves complex data structures. For example, the two-level RCT expression presented thus far has only represented the unconditional univariate outcome model. Causal inferencing via MLM or SEM rests on the inclusion of a group indicator to represent those individuals who were randomized to the treatment or control conditions. Figure 3a provides a

graphical display of a conditional multilevel model in a SEM framework where X_{ij} represents a dummy-coded variable to denote group assignment and κ is the prediction of Y_{ij} from X_{ij} . Figure 3b also displays a conditional multilevel model, in an *n*SEM framework, yet does so in a different way where there are now three levels (i.e., treatment group, control group, and classroom). Because *n*SEM is focused on groups of individuals and the *influences* within and across groups, the representation in graphical model displays the inherent structure of the data more readily: 1) there are students who were randomly assigned to the treatment group (level 1) who have a set of observed scores (Y_1^1) with an intercept (ν_1^1) and a residual variance ($\theta_{1,1}^{1,1}$), 2) there are students who were randomly assigned to the control group who have a set observed scores (Y_1^2) with an intercept (ν_1^2) and a residual variance ($\theta_{1,1}^{1,1}$), and 3) student outcomes in the treatment and control groups are both influenced by their classroom environments (level 3). Two comments on this particular structure are important to state; first, $\theta_{1,1}^{1,1}$ has been constrained to be equal across the student levels. This equality conforms to the expression of the conditional SEM in Figure 3a with θ . Second, the intercepts at levels 1 and 2 (ν_1^1 and ν_1^2) are freely estimated. The two equations that come from Figure 3b are

$$Y_1^1 = \nu_1^1 + \lambda_{1,1}^{1,3} \eta_1^3 + e_1^1 \quad \text{Eq. 4}$$

$$Y_1^2 = \nu_1^2 + \lambda_{1,1}^{2,3} \eta_1^3 + e_1^1 \quad \text{Eq. 5}$$

where η_1^3 are distributed $N[\alpha_1^3, \psi_{1,1}^{3,3}]$ and e_1^1 are distributed $N[0, \theta_{1,1}^{1,1}]$. By comparing the model in Figure 3b to a model where the intercepts in Eqs. 4 and 5 are constrained to be equal across student levels (i.e., ν_1^1), a log likelihood test will provide the test of mean differences in the outcome similar to the test expressed in Figure 3a.

Manifest Variable Tutorial: Fully Nested RCT (FN-RCT) with Students and Classrooms

At this point, it is useful to provide a tutorial for the example of students and classrooms used thus far as a way to demonstrate the similarities and differences in using multilevel and nSEM applications for nested data. The sample data of $n = 412$ participants used in this manuscript are drawn from a larger efficacy trial of a reading comprehension intervention (see Wanzek et al., 2017 for details on sampling, procedures, methods, and analyses). This study was exempt from IRB approval due to the use of de-identified data. The following files can be downloaded from the companion website:

1. *treatment.csv*: contains rows for $n = 199$ students who were randomly assigned to the treatment group and 11 columns for: treatment group ID (“treatment”), small group ID (“smallgroup”), classroom ID (“class”), group name (“treat”), group indicator (“passport”), student posttest score 1 (“postrc1”), student posttest score 2 (“postrc2”), student pretest score 1 (“prerc1”), student pretest score 2 (“prerc2”), grand mean centered pretest score 1 (“prerc1c”), and grand mean centered pretest score 2 (“prerc2c”).
2. *control.csv*: contains rows for $n = 213$ students who were randomly assigned to the control group and 11 columns for: control group ID (“control”), small group ID (“smallgroup”), classroom ID (“class”), group name (“treat”), group indicator (“passport”), student posttest score 1 (“postrc1”), student posttest score 2 (“postrc2”), student pretest score 1 (“prerc1”), student pretest score 2 (“prerc2”), grand mean centered pretest score 1 (“prerc1c”), and grand mean centered pretest score 2 (“prerc2c”).
3. *class.csv*: contains rows on $n = 50$ classrooms and two columns for: classroom ID (“class”) and the cluster size (“clustern”).

4. *smallgroup.csv*: contains data on $n = 43$ small groups and two columns for small group ID (“smallgroup”) and the cluster size (“clustern”).
5. *student.csv*: the combined data for *treatment.csv* and *control.csv*.

For all manifest variable examples, SAS PROC MIXED is used for the MLM representation of nesting and the *xxm* package (Mehta, 2013b) in R software is used for the *nSEM* representation. In each section we provide an introduction to the *nSEM* building process through *xxm* and a comparison to PROC MIXED. As each examples builds upon the previous in terms of structure, only the features of the model that are unique to subsequent examples will be provided in text for the tutorial. Full, annotated code for each example is provided in the supplemental online materials.

Example 1: Multilevel and *nSEM* Comparison of Unconditional Models. To illustrate the univariate unconditional model from Figure 1 using MLM (Figure 1a) and *nSEM* (Figure 2b), the *fulldata.csv* file is used for both *xxm* and PROC MIXED. *nSEM* specification through *xxm* requires the user to specify the *model* (i.e., the number of levels), one to n *submodels* (i.e., level-to-level relations with identified outcomes and predictors), and *within-* and *between-*level matrices. It is further necessary to create separate files for each level identified in the *model*. For this first example of students and classrooms, the *student.csv* and *class.csv* should be loaded into R along with the *xxm* package. The first command for the univariate unconditional model is

```
ex1<-xxmModel(levels=c("student","class")).
```

The object *ex1* creates an *xxm* model with two levels, student and class. The *c()* function combines the student and class vectors that are created next with *submodel* commands

```
ex1<-xxmSubmodel (model=ex1,level="student", parents="class", ys="postrc1", xs=, etas=,
data=studentr1).
```

The first submodel defined above is for the student (`level="student"`) and the nesting of students in classes is defined by `parents="class"`. The outcome for this example, `postrc1`, is defined by `ys="postrc1"`, predictors can be listed next to `xs=`, the `etas=` command defines latent variables (not present at the student level), and the `data=studentr1` command names the student data frame. Although each command within the Submodel does not need to have listings (e.g., there are no predictors or latent variables at the student level), all seven commands must appear in the code. Next, the classroom submodel can be added with

```
ex1<-xxmSubmodel(model=ex1, level="class", parents=, ys=, xs=, etas="intercept", data=classr1).
```

The difference between the two submodels is that the classroom model has no parent (i.e., classrooms are not nested within another level), and the outcome related to `postrc1` in the student submodel is now expressed as `etas="intercept"`.

Once the model and submodels have been defined, four matrices are added that correspond to Figure 2b: the latent mean (α_1^2), the latent variance ($\psi_{1,1}^{2,2}$), the cross-level $\lambda = 1.0$ linking of η_1^2 to Y_1^1 , and the residual variance ($\theta_{1,1}^{1,1}$). Each matrix is programmed with two components: 1) a pattern matrix that defines whether the parameter is fixed (represented by a “0”) or freely estimated (represented by a “1”), as well as the number of rows and columns, and 2) a value matrix that provides the start value at the row x column portion of the matrix. For example, the latent mean is specified with

```
alpha_pattern<-matrix(c(1),1,1)
alpha_value<-matrix(c(457),1,1)
```

where the pattern matrix defines that 1 mean is freely estimated `c(1)` in a 1 x 1 matrix `(1,1)`. The alpha value matrix specifies a starting value of 457 for the latent mean in the 1 x 1 alpha matrix. Next, the latent variance matrix is added

```
psi_pattern<-matrix(c(1),1,1)
```



```
psi_value<-matrix(c(2.6),1,1)
```

where 2.6 is a starting value for the freely estimated $\psi_{1,1}^{2,2}$, and the $\theta_{1,1}^{1,1}$ matrix is added in a similar manner with

```
theta_pattern<-matrix(c(1),1,1)
theta_value<-matrix(c(467),1,1).
```

The cross-level linking of η_1^2 to Y_1^1 via λ occurs as

```
lambda_pattern<-matrix(c(0),1,1)
lambda_value<-matrix(c(1),1,1).
```

Notice that the pattern matrix specifies that λ should be fixed and the value matrix gives a value of 1.0 in accordance with Figure 2b.

The final step in `xxm` is to connect the pattern and value matrices to the levels. Of the four parameter-type matrices specified, two are within-classroom matrices (i.e., alpha and psi), one is a within-student matrix (i.e., theta), and one is a cross-level matrix (i.e., lambda).

Subsequently, four commands are given to connect the value and pattern matrices to levels. The connection of alpha and psi matrices to the classroom are made by

```
ex1<-xxmWithinMatrix (model=ex1, level="class", type="alpha", pattern=alpha_pattern,
value=alpha_value)
ex1<-xxmWithinMatrix (model=ex1, level="class", type="psi", pattern=psi_pattern, value=psi_value)
```

where `xxmWithinMatrix` defines the type of matrix, `model=ex1` defines the model with which the matrix is associated, `level="class"` states the level at which the matrix is associated, `type=` states the type of matrix that is being connect to the level (i.e., alpha and psi), and then the `pattern=` and `value=` commands denote the respective pattern and values matrices that were previously programmed. The connection of the theta matrix to the student level is made by

```
ex1<-xxmWithinMatrix (model=ex1, level="student", type="theta", pattern= theta_pattern, value=
theta_value)
```

and the link between the classroom and student (i.e., cross-level matrix) is added with

```
ex1<-xxmBetweenMatrix (model=ex1, parent="class", child="student", type="lambda",
  pattern=lambda_pattern, value=lambda_value)
```

where both the parent and child levels must be defined in order to correctly link the classroom and student levels.

With all of the matrices defined and connect, the *nSEM* can be estimated with the following three commands

```
ex1<-xxmRun(ex1)
xxmCI(model=ex1)
xxmSummary(model=ex1)
```

where the first command estimates the model *ex1*, the second command estimates confidence intervals for the results, and the third command summarizes the first two and includes the number of parameters, log likelihood, the sample size, the AIC and the BIC for the model.

Results from this example are reported in Table 1 alongside findings from using PROC MIXED (see supplemental online materials). The fixed and random effect coefficients from the multilevel and *nSEM* approaches are identical for the intercept (463.23), classroom variance (99.11), and residual variance (566.98). Moreover, the confidence intervals across methods approximated each other well for each effect.

Example 2: Multilevel and *nSEM* Comparison of Causal Models. Moving from the basic illustration of the unconditional means model, we now provide an illustration of causal modeling. Here, the direct congruence between MLM and *nSEM* approaches deviate as shown in Figures 3a and 3b, respectively. To estimate the causal model for the full nesting of students in classrooms, we now use the *treatment.csv*, *control.csv*, and *classroom.csv* data files and load them into R. From an *nSEM* programming perspective in *xxm* and using Figure 3b as a guide, we may observe that there will be three levels, two within-level pattern and value matrices at the

classroom level for α_1^3 and $\psi_{1,1}^{3,3}$, and a within-level pattern and value matrices for $\theta_{1,1}^{1,1}$ for each of the treatment and control levels. In addition to these components, we now add a nu matrix for the ν_1^1 and ν_1^2 observed means estimated at the treatment and control levels, respectively. For the configuration in Figure 3b, the xxm begins with

```
ex2<-xxmModel(levels=c("treatment","control","class"))
```

where the three levels of treatment group, control group, and classrooms are defined. The submodels associated with each level are

```
ex2<-xxmSubmodel(model=ex2,level="treatment", parents="class", ys="postrcl", xs=etas,
  data=treatmentr1)
ex2<-xxmSubmodel(model=ex2,level="control", parents="class", ys="postrcl", xs=etas,
  data=controlr1)
ex2<-xxmSubmodel(model=ex2,level="class", parents=,ys=, xs=, etas=c("int"),data=classr1)
```

where the treatment and control submodels show that they are both nested within classrooms.

The pattern and value matrices are the same for psi, theta, and lambda as in Example 1, save for an addition to theta matrix programming. Labels in xxm serve to provide equality constraints across groups for specified parameters. In the current example, we assume that $\theta_{1,1}^{1,1}$ is the same for both the treatment and control groups, and by specifying the following command, the construction of the within-group matrices will apply the constraint during estimation:

```
theta_label<-matrix(c("thetac"),1,1).
```

The treatment and control intercepts are added via the defined nu pattern and value matrices

```
nu_pattern<-matrix(c(1),1,1)
nu_value<-matrix(c(457),1,1).
```

As with the previous examples, the pattern matrix here states that nu should be freely estimated with a starting value of 457 in the value matrix. The final change to the matrix programming pertains to the alpha matrix. Unlike the first example, where the latent mean was estimated at the classroom level for all students, the means are now estimated via the nu matrices. As such, the

alpha matrices are now represented with the following to designate that a mean is not to be estimated at the classroom level.

```
alpha_pattern<-matrix(c(0),1,1)
alpha_value<-matrix(c(0),1,1)
```

Connecting the matrices to levels maintains an identical process as in Example 1. The alpha and psi matrices are connected to the classroom example using the same code as Example 1 but changing the object and model names to “ex2”. At the treatment and control group levels, respectively, the nu and theta matrices are added in the same way that student matrices were added in Example 1:

```
ex2<-xxmWithinMatrix(model=ex2,level="treatment", type="nu", pattern=nu_pattern, value=nu_value)
ex2<-xxmWithinMatrix(model=ex2,level="treatment", type="theta", pattern=theta_pattern,
  value=theta_value, label=theta_label)

ex2<-xxmWithinMatrix(model=ex2,level="control", type="nu", pattern=nu_pattern, value=nu_value)
ex2<-xxmWithinMatrix(model=ex2,level="control", type="theta", pattern=theta_pattern,
  value=theta_value, label=theta_label).
```

Note that the theta matrices now include the label command, constraining the coefficient to equality during estimation. The construction of the between-matrix requires to commands, one for each of the treatment and control groups. As Figure 3b shows separate cross-level links from η_1^3 to Y_1^1 and Y_1^2 , separate between-level matrices are specified

```
ex2<-xxmBetweenMatrix(model=ex2,parent="class",child="treatment", type="lambda",
  pattern=lambda_pattern, value=lambda_value)
ex2<-xxmBetweenMatrix(model=ex2,parent="class",child="control",type="lambda",
  pattern=lambda_pattern, value=lambda_value).
```

With all matrices constructed and linked, the model may be estimated as before. Results for the nSEM causal model of students nested within classrooms are compared to the multilevel model and reported in Table 1. Both classroom and residual variance coefficients are identical across

methodologies, with nearly identical confidence intervals. The primary difference in reporting can be seen in the fixed effects. The multilevel representation of the causal model is

$$Y_{ij} = \gamma_{00} + \gamma_{01}(Treatment)_{ij} + u_{0j} + e_{ij}$$

such that the resulting coefficients in Table 1 are expressed through the intercept and slope. Because $\gamma_{01}(Treatment)_{ij}$ is a dichotomous variable, the intercept of 457.22 represents the mean of the control group, and the slope value of 12.23 is the positive, fitted deflection from the mean. As seen in Figure 3b and Eqs. 4 and 5 for this model, the nSEM approach estimates the marginal means directly for each level. The means for in Table 2 both reflect this estimation and provides congruence with the multilevel model, such that the difference between the Control Mean (457.22) and the Passport Mean (469.45) is equal to the Passport coefficient in the multilevel model (i.e., 12.23). The statistical test of equality between ν_1^1 and ν_1^2 is evaluated by a re-estimation of the nSEM model but including an equality constraint in the nu matrix such as

```
nu_label<-matrix(c("txcon_nu"),1,1)
```

and performing a log likelihood test. In the present example the difference between the freed (deviance = 3730.51) and constrained (3757.13) models was statistically significant ($\Delta\chi^2 = 26.62$, $\Delta df = 1$, $p < .001$) favoring the freed model and pointing to a significant treatment effect.

Manifest Variable Tutorial: Partial Nested RCT (PN-RCT) with Students and Small Groups

Suppose now that the classroom influence on students is ignored and we consider the partial nesting influence on treatment students. An example graphical illustration for this model is presented in Figure 4a. Note that even though Figure 3b and 4a both present configurations of a three-level model, they differ whereby the former includes the third level for the full nesting of

students in classrooms and the latter includes the third level to characterize that only treatment students are influenced by small groups. The equations associated with this model are

$$Y_1^1 = \nu_1^1 + \lambda_{1,1}^{1,3} \eta_1^3 + e_1^1 \quad \text{Eq. 6}$$

$$Y_1^2 = \nu_1^2 + e_1^1 \quad \text{Eq. 7}$$

and denote that the only the treatment group maintains an influence of small groups on the outcome. xxm software is flexible to estimate this model and is done so in a manner similar to previous examples.

Example 3: Multilevel and *n*SEM Comparison of Causal Models. The specification of models, submodels, and matrices for the partially nested condition deviates very minimally from the fully nested, causal model in Example 2. The xxm model associated with Figure 4a is

```
ex3<-xxmModel(levels=c("treatment","control","smallgroup"))
```

showing that we have the three levels of treatment, control, and small groups. The submodel commands for each level are as follows

```
ex3<-xxmSubmodel(model=ex3,level="treatment",parents="smallgroup",ys="posttest",xs=,etas=,
  data=treatmentr1)
ex3<-xxmSubmodel(model=ex3,level="control",parents=,ys="posttest",xs=,etas=,data=controlr1)
ex3<-xxmSubmodel(model=ex3,level="smallgroup",parents=,ys=,xs=,etas=c("int"),data=smallgroup1)
```

where the small group submodel is identical in form and function as the classroom model in the previous example, and treatment students are nested within small groups. The primary difference between the partial and fully nested submodel code is in the control group where there is no affiliation with an upper-level unit. In this way, the expectation for matrix programming is that only the between-matrix linking of levels will change. The pattern and value programming of alpha, psi, nu, lambda, and theta matrices are identical to Example 2, as are the treatment and control group's within-level nu and theta linking to those matrices. Whereas Example 2 demonstrated that two between-level matrices were needed to separately link the classrooms to

treatment and small groups, the partial nesting necessitates only one between-level matrix connecting the small groups to the treatment group

```
ex3<-xxmBetweenMatrix(model=ex3,parent="smallgroup",child="treatment", type="lambda",
  pattern=lambda_pattern, value=lambda_value).
```

The model is then estimated with the same three commands as before with results for provided in Table 1. The estimates of the residual variance are the same across method (multilevel = 549.16, *nSEM* = 549.18) as are the small group variances (multilevel = 127.60, *nSEM* = 127.59). The intercept from the multilevel model is identical to Control group intercept (i.e., 455.49), and the difference between the Control and Treatment group intercepts is identical to the Passport slope coefficient in the multilevel model (i.e., 12.03). The confidence intervals for the Control intercept and residual variance are approximately similar across methodologies, with a wider estimated range interval of small group variance observed when using the multilevel approach (95% CI: 63.45, 375.84) compared to the *nSEM* (95% CI: 44.36, 276.74). A possible reason for the difference in confidence interval estimates for the small group variance may lie in the methodology system itself. Causal inferencing with traditional mixed modeling packages that assume full nesting (e.g., SAS and Mplus) require data manipulations, additional clustering considerations in programming, and adjustments to estimation that are not specifically inherent to the data. It is possible that these data manipulations, especially when pertaining to required estimation changes (e.g., the need to use `ddfm = satch` in PROC MIXED), may lead to biased confidence interval estimates that would not be otherwise observed when the natural clustering of the data is modeled without the need for additional data adjustments prior to testing, such as when using *nSEM*.

Manifest Variable Tutorial: Partial Nested, Cross-Classified Model Causal Model

The final manifest variable examples combine the full nesting context of students in classrooms and the partial nesting of treatment group students in small groups (Figure 4b). The combined influences of classrooms and small groups are now represented by levels 3 and 4, respectively, in the model and are observed in scalar form as

$$Y_1^1 = \nu_1^1 + \lambda_{1,1}^{1,3}\eta_1^3 + \lambda_{1,1}^{1,4}\eta_1^4 + e_1^1 \quad \text{Eq. 8}$$

$$Y_1^2 = \nu_1^2 + \lambda_{1,1}^{2,3}\eta_1^3 + e_1^1 \quad \text{Eq. 9}$$

where η_1^4 is distributed $N[\alpha_1^4, \psi_{1,1}^{4,4}]$ and the other parameters as distributed as previously reported. The combination of cross-classification, full nesting in classrooms, and partial nesting in small groups is flexibly presented in both equations and measurement models. Rather than a change to the model subscript notation as in MLM and SEM, the *n*SEM equations in Eqs. 8 and 9 are only distinguished by addition of η_1^4 with the associated cross-level λ links that are fixed at 1.0. The test of causal inference is identical to that presented in example 2 where the intercepts in Eqs. 6 and 7 are differentially constrained to equality and then freed with a log likelihood test of model differences.

Example 4: Comparison of Partially Nested Causal Models. By combining elements of Examples 2 and 3, the *n*SEM analysis in xxm becomes quite simple, even for a four-level, multilevel, cross-classified, partially nested randomized trial. Because the configuration of Figure 4b contains elements of the fully nested example in Figure 3b and the partially nested example in Figure 4a, the implications for the matrices in Figure 4b are that the treatment group should have two within-level matrices (nu and theta), the control group should have two within-level matrices (nu and theta), the classrooms should have two-within level matrices (alpha and psi), the small groups should have two within-level matrices (alpha and psi), and then three

between-level matrices should be coded to account for the influence of classrooms on the treatment group, classrooms on the control group, and small groups on the treatment group. The model statement associated with Figure 4b is

```
ex4<-xxmModel(levels=c("treatment","control","class","smallgroup")).
```

The submodel for the control group is identical to the statement from Example 1 where the parent is the classroom. The class and subgroup submodels are also nearly identical to previous iterations and are only altered by the need to give different labels for the etas (e.g., “int” for class and “int1” for smallgroup). Only the treatment group differs here due to the presence of both classroom and small group influences

```
ex4<-xxmSubmodel(model=ex4,level="treatment", parents=c("class","smallgroup"), ys="postrcl", xs=,
  etas=, data=treatmentr1).
```

The alpha, psi, nu, lambda, and theta matrices are all written identically to those presented in Example 3 and the within-level matrices for each level are also written identically to those presented in Example 3. Even the construction of the between-level matrices are identical when linking the classrooms to treatment and control groups (Example 2) and the small groups to the treatment group (Example 3). Once the full model has been built, the model may be estimated; results are reported in Table 1. Again we observe symmetry between using a traditional multilevel approach with *n*SEM for the intercept and control intercept coefficients (456.85), the classroom variance (70.77 in the multilevel and 70.87 in *n*SEM), and the residual variance (518.67). Only in the small group variance and confidence intervals do we observe differences between the multilevel (43.25, 95% CI = 15.40, 1,348.64) and *n*SEM (53.21, 95% CI = .001, 184.74) approaches.

Latent Variable Models

As frequent as causal modeling is with observed outcomes, there exists a parallel call for increased focused on measurement at the construct level. Because of inherent limitations to

observed outcome models (e.g., poor precision, truncated estimates of effect sizes), latent variable modeling serves as an effective solution to such limitations for both causal modeling and the study of individual differences. Suppose that Example 4 is now extended in two ways: 1) multiple, observed measures of reading comprehension are used to form a common factor of reading comprehension, and 2) baseline performance of latent reading comprehension is used as a covariate. Figure 5 provides a picture of this model whereby the four-level model of treatment students, control students, classrooms, and small groups are included, as well as latent post-test constructs of reading comprehension (i.e., η_1^1 and η_1^2), latent baseline constructs of reading comprehension (i.e., η_2^1 and η_2^2), factor loadings of the observed variables on the latent variables (i.e., $\lambda_{1,1}^{1,1}$ and $\lambda_{2,1}^{1,1}$), and the regression of latent post-test on the latent baseline construct (i.e., $\beta_{1,2}^{1,1}$), where $\beta_{p,q}^{u,v}$ describes the relation of the q th latent variable at level v , to the p th latent variable at level u . Note that Figure 5 includes five constraints that are part of the model building and testing process: 1) $\beta_{1,2}^{1,1}$ is constrained to be equal across levels 1 and 2 and allows for an explicit test of the homogeneity of regression (i.e., baseline moderation of treatment effects), 2) $\lambda_{1,1}^{1,1}$ is constrained across Y_1^1, Y_3^1, Y_1^2 , and Y_3^2 to assume measurement invariance of the first reading comprehension measure across time points (baseline and post-test) and levels (treatment and control), 3) $\lambda_{2,1}^{1,1}$ is constrained across Y_2^1, Y_4^1, Y_2^2 , and Y_4^2 to assume measurement invariance of the second reading comprehension measure across both time points and levels, 4) $\theta_{1,1}^{1,1}$ is constrained across Y_1^1, Y_3^1, Y_1^2 , and Y_3^2 to assume measurement invariance for the residual of the first reading comprehension measure across time points and levels, and 5) $\theta_{2,1}^{1,1}$ is constrained across Y_2^1, Y_4^1, Y_2^2 , and Y_4^2 to assume measurement invariance for the residual of the second

reading comprehension measure across time points and levels. All such constraints can empirically tested to evaluate at which places strict or partial measurement invariance exists for the given parameters.

When considering the equations for the latent variable version of Example 4, the measurement model for treatment group students' reading comprehension achievement can be written as:

$$Y_1^1 = \nu_1^1 + \lambda_{1,1}^{1,1}\eta_1^1 + e_1^1$$

$$Y_2^1 = \nu_2^1 + \lambda_{2,1}^{1,1}\eta_1^1 + e_2^1$$

$$Y_3^1 = \nu_3^1 + \lambda_{1,1}^{1,1}\eta_2^1 + e_1^1$$

$$Y_4^1 = \nu_4^1 + \lambda_{2,1}^{1,1}\eta_2^1 + e_2^1$$

where sets of $\lambda_{1,1}^{1,1}$ and $\lambda_{2,1}^{1,1}$ as well as e_1^1 and e_2^1 are differentially constrained across the observed measures in accordance with assumption of measurement invariance loadings residuals, the intercepts are freed across all observed measures ($\nu_1^1 - \nu_4^1$), e_1^1 is distributed $N[0, \theta_{1,1}^{1,1}]$, e_2^1 is distributed $N[0, \theta_{2,1}^{1,1}]$, and η_2^1 is distributed $N[0, \psi_{2,2}^{1,1}]$. The structural model for the regression of latent post-test on latent baseline is then

$$\eta_1^1 = \lambda_{1,1}^{1,3}\eta_1^3 + \lambda_{1,1}^{1,4}\eta_1^4 + \beta_{1,2}^{1,1}\eta_2^1 + \xi_1^1$$

where the factor loadings $\lambda_{1,1}^{1,3}$ and $\lambda_{1,1}^{1,4}$ are fixed at 1.0 linking the small groups and classrooms, respectively, to the treatment group, η_1^3 and η_1^4 are distributed as previously shown, $\beta_{1,2}^{1,1}\eta_2^1$ is the regression parameter, and ξ_1^1 is distributed $N[\alpha_1^1, \psi_{1,1}^{1,1}]$. The control group equations are nearly

identical to those presented for the treatment group with the exception that there is no linking of η_1^3 to η_1^2 , and η_u^p are distributed with corresponding changes to variable number and level.

Example 5: Comparison of Latent Partially Nested Causal Models. Where our final example departs from the previous examples are the introduction of the latent variables as well as the inclusion of baseline achievement data. These additions necessitate extra steps in model building such as longitudinal invariance testing within- and between- levels as well as individual time-point invariance testing between levels. Due to the way that *n*SEM structures the model and the number of invariance tests, we have opted to include five sets of invariance testing in the supplemental online materials that can be applied to the sample data. The specification of the latent partially nested model in *xxm* requires additional considerations in the construction of the matrices pertaining to the scaling of the latent variable and expanded matrices to accommodate the inclusion of multiple indicators across multiple latent constructs. The creation of the submodels are parallel to previous iterations and are extended to include the baseline measures. For example, the level one submodel for the treatment condition is

```
impact<-xxmSubmodel(model=impact,level="treatment", parents=c("smallgroup","class"),
  ys=c("postrc1","postrc2","prerc1","prerc2"),xs=,etas=c("spring","fall"),data=treatmentr1)
```

where the *ys* now include two indicators for latent reading comprehension at the post-test (i.e., *eta*="spring") and two indicators for baseline latent reading comprehension (i.e., *eta*="fall").

Different from the manifest variable examples, separate pattern and value matrices are created for levels 1 and 2 for α in order to reflect an estimated baseline alpha for the treatment group (see *txalpha* in supplemental online materials) but fixed at zero in the control group (see *conalpha* in supplemental online materials) for longitudinal scaling and providing standardized means.

Similarly, the ψ matrix for the treatment group allows the baseline variance to be freely estimated but fixed at 1 in the control group, again, for developmental scaling and standardized

means. In totality, the inclusion of four indicators in each of the treatment and groups necessitates the following within-level matrices for the treatment and control groups:

- 2x1 α matrix
- 2x2 ψ matrix
- 4x4 θ matrix with variances estimated on the diagonal and covariances on the off-diagonal
- 4x2 λ matrix for the loadings of manifest variables on the level-1 and level-2 latent variables
- 2x2 within-level β matrix for the regression of latent post-test on baseline
- 4x1 v matrix.

The classroom and small group levels include the α and ψ matrices as in previous examples, as well a 2x1 between-level β matrix, with values fixed at 1.0, for the linking of small groups and classes to the latent post-tests in the treatment and control groups. Once the model is specified with appropriate constraints, a sample result is included in Figure 6. Findings include a standardized latent post-test mean of 0.85 for the control group along with a standardized value of 1.73 for the treatment group. The fit for this model with freed intercepts for the treatment and control was $-2LL = 13,199.94$ with 17 parameters, $AIC = 13,233$, $BIC = 13,326$. When compared to a model that constrained the intercepts to be equal across levels 1 and 2, the fit degraded to $-2LL = 13,229.98$ with 16 parameters, $AIC = 13,262$, $BIC = 13,348$. The difference in deviance statistics was statistically meaningful ($\Delta-2LL = 30.04$, $\Delta df = 1$, $p < .001$) and because the level 1 and level 2 means were standardized values, the difference represents the average treatment effect (i.e., $1.73 - 0.85 = 0.88$).

Discussion

There exists an abundance of resources for modeling complex data structures in terms of measurement frameworks (e.g., MLM and SEM) and software. The relation between complex data structures in the context of randomized controlled trials has been studied extensively such that comprehensive materials and best practice recommendations exist for the prospective planning of randomized studies (Spybrook, Raudenbush, Liu, Congdon, & Martinez, 2006) and subsequent evaluation of causal impacts (What Works Clearinghouse, 2016). These strategies and their predecessors have proven useful for decades of research so that unbiased estimates of treatment effects may be estimated. At the same time, a noted limitation of existing frameworks, strategies, and best practice recommendations is that they have been focused on fully nested design structures.

Partial nested designs have gained greater traction in psychological research (Sterba et al., 2014), yet even as these designs are widely utilized for the implementation of treatment effect studies in educational and psychological research, only recently have methodological recommendations begun to emerge for handling such structures in terms of both theoretical considerations for the model (e.g., Bauer et al., 2008; Lee & Thompson, 2005) and practical application for sample size planning and estimation (e.g., Lohr et al., 2014; Sterba et al., 2014). When modeling nested data, whether partially nested, fully nested, or otherwise, most applications have focused on multilevel or structural equation modeling frameworks for estimation (Sterba et al., 2014), and the current recommendations require the user to engage in additional data management and variable creation in order to override or allow the model to be estimated in the appropriate manner. In the present study, our goal was to both generally introduce *n*SEM as a viable framework for handling design complexities of many kinds of nesting structures and specifically introduce *n*SEM as useful for handling partially nested data

for causal modeling. The development of *nSEM* modeling and software by Mehta (2013a, 2013b) was born out of an attempt to bridge MLM, SEM, and ML-SEM traditions into a general unifying framework that allows for direct correspondence among matrices, measurement models, and equations. As well, *nSEM* was developed as a framework to flexibly handle many levels of structure, whether the levels represent the sources of influence (e.g., higher-order clustering units, cross-classification, multiple membership) as they do in an MLM framework, or categorical grouping variables (e.g., treatment conditions, sex, or groupings of socio-economic status) as they are handled in a SEM framework.

The answer to our research question of, *to what extent does an nSEM framework yield identical or similar results to conventional applications of partial nesting for a randomized controlled trial?*, is that in the context of the sample data results from the examples on manifest variables demonstrated a strong correspondence between *nSEM* and MLM approaches. However, it is important to note the differences in model conceptualization and estimation. Though producing the same result as *nSEM*, the MLM specification requires additional data management and code specification to get around estimation issues. Specifically, MLM applications require the creation of multiple variables, addition of controlling statements, and changes in estimation procedures that are not typically part of estimating mixed models in software packages (Lohr et al., 2014). Where *nSEM* differs from MLM applications in conventional software packages is that its straightforward modeling of influences allows for a more direct model building approach. A conventional two-level model for randomized controlled trials in MLM (e.g., students in classrooms) can instead be viewed as a three-level model in *nSEM* (i.e., treatment students, control students, and classrooms). The estimation of the treatment effect in the MLM is typically obtained by including a dummy-coded indicator that

reflects one group's fitted post-test mean deflection from the other group, whereas in *n*SEM the treatment effect comes from testing and comparing a model with freely estimated group means from a model of constrained group means. Notwithstanding these differences, the results are just about identical except for differences in the estimated confidence intervals for the small group variance, a potential consequence of including extra data manipulations and estimation procedures in the MLM approach. The *n*SEM for ML-SEM applications differs from what a traditional ML-SEM approach in specification in that the former views the causal model of as a four-level model (i.e., treatment students, control students, small groups, and classrooms) and the latter views the structure as three-level with the two student groups collapsed. The specification of multiple sources of influence in *n*SEM allows for a broad set of longitudinal invariance models to be specified and tested across all four levels in a way that more directly models the individual parameters that exist across any level.

Importantly, even as each example demonstrates the unique value of *n*SEM, Example 5 adds unique value that addresses limitations of the partial nesting literature raised by Sterba et al. (2014); namely, that when considering partial nesting a balance is needed in decision-making about the type of model, the nature of form of the data (long or wide), and how specialized the software package is for running multilevel or SEMs. Findings from this tutorial demonstrate that regardless of variable type (e.g., manifest or latent), nesting design (full or partial), leveling of the model (single level or multilevel), or complexity of the design structure (e.g., cross-classification), the *n*SEM approaches assists for what other data management mechanisms may ail. With relatively little training in *n*SEM materials (Mehta, 2013b) it is possible to specify the types of complex models here as well as alternatives that are frequently germane to causal modeling in psychological research. For example, Wanzek et al. (2017) expanded the use of

*n*SEM to test for moderation of treatment effects through: 1) heterogeneity of regression slopes, 2) the influence of continuous predictors on treatment effects, and 3) the influence of categorical predictors on treatment effects. Heterogeneity of regression slopes is commonly tested in MLM and SEM causal models via an interaction term of the treatment indicator and baseline performance. In *n*SEM, heterogeneity is tested by freeing the latent posttest on pretest regression parameter, such as $\beta_{1,2}^{1,1}$ in Figure 5, and comparing the model fit for the presence of varying regression effects across groups. The influence of continuous predictors may be tested by creating additional level-based latent constructs, whether common-factor or single-indicator, and testing freely estimated effects compared to constrained effects. The influence of categorical predictors may be tested by creating additional levels or subsetting the data pertaining to the levels of interest and testing for treatment effects for the subgroups of interest. Multilevel moderation and mediation is also possible in *n*SEM through xxm software (Mehta, 2013b).

Conclusion

Complex data structures are here to stay. Methodologies for understanding and estimating such complexities continue to grow, and as such, approaches will continually be refined for the purpose of estimating treatment effects in ways that are unbiased and useful to the practitioner and the data modeler. Although the *n*SEM approach introduces added complexities in equation notation, the framework allows for model specification malleability that is only as limited as one's computer resources and data allows.

References

- Baldwin, S. A., Bauer, D. J., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. *Psychological Methods, 16*, 149. Doi: [10.1037/a0023464](https://doi.org/10.1037/a0023464).
- Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research, 43*, 210-236. Doi: [10.1080/00273170802034810](https://doi.org/10.1080/00273170802034810).
- Curran, P. J. (2003). Have multilevel models been structural equation models all along?. *Multivariate Behavioral Research, 38*, 529-569. Doi: [10.1207/s15327906mbr3804_5](https://doi.org/10.1207/s15327906mbr3804_5).
- Gilbert, J., Petscher, Y., Compton, D. L., & Schatschneider, C. (2016). Consequences of misspecifying levels of variance in cross-classified longitudinal data structures. *Frontiers in psychology, 7*, 695.
- Lee, K.J., & Thompson, S.G. (2005). The use of random effects models to allow for clustering in individually randomized trials. *Clinical Trials, 2*, 163–173. Doi: [10.1191/1740774505cn082oa](https://doi.org/10.1191/1740774505cn082oa).
- Luo, W., Cappaert, K. J., & Ning, L. (2015). Modelling partially cross-classified multilevel data. *British Journal of Mathematical and Statistical Psychology, 68*(2), 342-362.
- Luo, W., & Kwok, O. M. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research, 44*(2), 182-212.
- Mehta, P.D. (2013a). N-level structural equation modeling. In Y. Petscher, C. Schatschneider, and D. Compton (Eds.) *Applied quantitative analysis in education and the social sciences* (pg. 329-361). NY, NY: Routledge.

- Mehta, P.D. (2013b). *xm: Structural Equation Modeling for Dependent Data*. R package version 0.6.0.
- Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*, 259. Doi: 10.1037/1082-989X.10.3.259.
- Mehta, P.D., & Petscher, Y. (2016). N-level structural equation model of student achievement data nested with repeated teachers, schools, and districts. In J.R. Harring, L. M. Stapleton, and S.N. Beretvas (Eds.). *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pg. 193-228). NY, NY: Information Age Publishing.
- Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41*(4), 473-497.
- Moerbeek, M., & Wong, W. (2008). Sample size formulae for trials comparing group and individual treatments in multilevel model. *Statistics in Medicine, 27*, 2850-2864.
- Roberts, C. & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials, 2*, 152–162.
- Sanders, E. A. (2011). *Multilevel analysis methods for partially nested cluster randomized trials*. University of Washington.
- Spybrook, J., Raudenbush, S. W., Liu, X. F., Congdon, R., & Martínez, A. (2006). *Optimal design for longitudinal and multilevel research: Documentation for the “Optimal Design” software*. Ann Arbor: University of Michigan School of Education, Hierarchical Models Project.

- Sterba, S.K., Preacher, K.J., Forehand, R., Hardcastle, E.J., Cole, D.A., & Compas, B.E. (2014). Structural equation modeling approaches for analyzing partially nested data. *Multivariate Behavioral Research, 49*, 93-118.
- Sterba, S.K. (2017) Partially nested designs in psychotherapy trials: A review of modeling developments, *Psychotherapy Research, 27, 4*, 425-436.
- Wanzek, J., Petscher, Y., Al Otaiba, S., Rivas, B., Jones, F., Kent, S., Schatschneider, C. & Mehta, P. (2017). Effects of a year-long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology*. Doi: 10.1037/edu0000184.
- What Works Clearinghouse (2016). *Procedures and standards handbook* (Version 3.0). Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

Table 1

Manifest variable example results

Example	Method	Effect	Estimate	95% CI	
				LB	UB
Example 1	Multilevel	Intercept	463.23	459.30	467.16
		Classroom Variance	99.11	51.33	265.72
		Residual Variance	566.97	491.77	660.95
	nSEM	Intercept	463.23	459.38	467.31
		Classroom Variance	99.11	38.20	204.42
		Residual Variance	566.98	490.90	659.63
Example 2	Multilevel	Intercept	457.22	452.71	461.73
		Passport	12.23	7.65	16.81
		Classroom Variance	98.63	51.96	254.57
		Residual Variance	528.68	458.58	616.26
	nSEM	Control Mean	457.22	452.81	461.80
		Passport Mean	469.45	465.02	474.03
		Classroom Variance	98.63	39.81	200.02
		Residual Variance	528.67	457.77	615.03
Example 3	Multilevel	Intercept	455.49	452.28	458.69
		Passport	11.95	6.11	17.79
		Small Group Variance	127.60	63.45	375.84
		Residual Variance	549.16	477.30	638.66
	nSEM	Control Mean	455.49	452.29	458.69
		Passport Mean	467.44	462.49	472.31
		Small Group Variance	127.59	44.36	276.74
		Residual Variance	549.18	476.47	634.40
Example 4	Multilevel	Intercept	456.85	452.64	461.07
		Passport	12.02	6.86	17.19
		Classroom Variance	70.77	29.99	321.69
		Small Group Variance	43.25	15.40	1348.64
		Residual Variance	518.67	448.46	606.86
	nSEM	Control Mean	456.85	452.73	461.28
		Passport Mean	468.88	463.93	473.76
		Classroom Variance	70.87	10.21	174.29
		Small Group Variance	53.21	0.001	187.74
		Residual Variance	518.67	447.49	605.37

Note. Example 1 = Unconditional means model with students and classrooms, Example 2 = Conditional means model with students and classrooms, Example 3 = Conditional means model with students and small groups, Example 4 = Conditional means model with students, classrooms, and small groups. nSEM = n-level structural equation model; LB = 95% confidence interval (CI) lower bound estimate; UB = 95% confidence interval (CI) upper bound estimate.

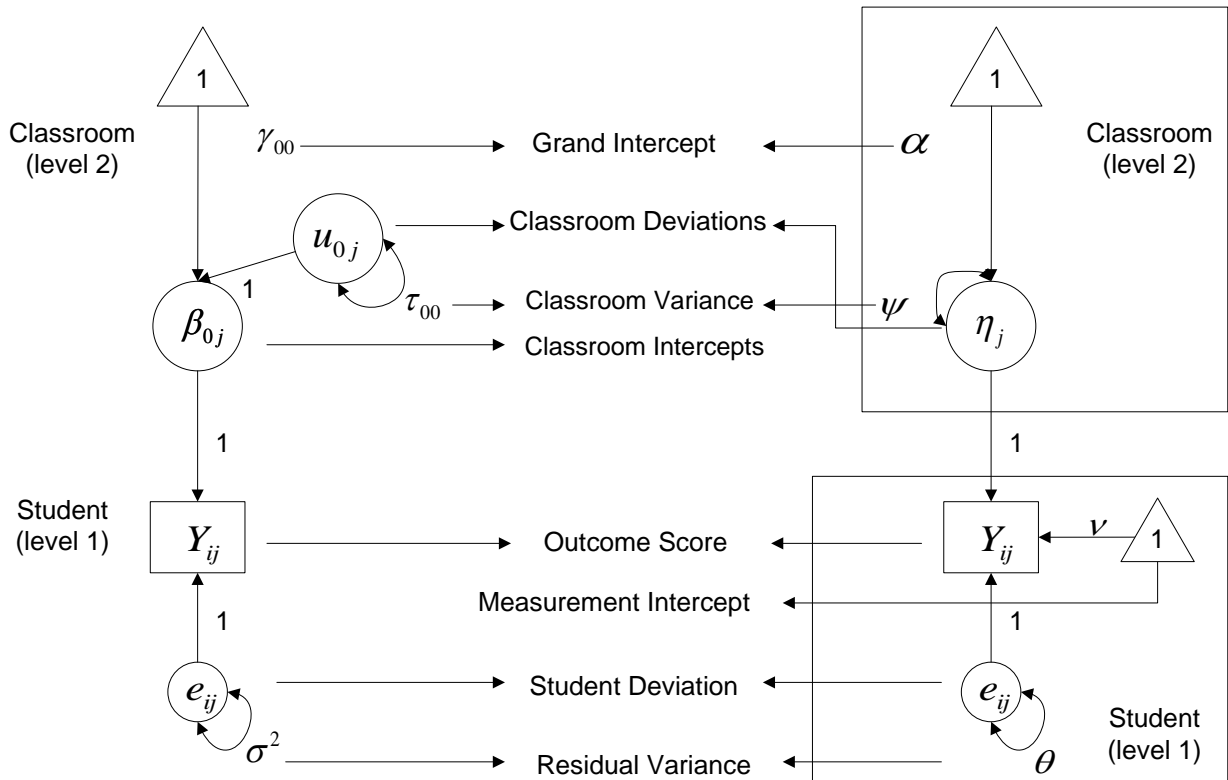


Figure 1. Diagram of a two-level model using hierarchical linear modeling notation (left) and structural equation modeling notation (right).

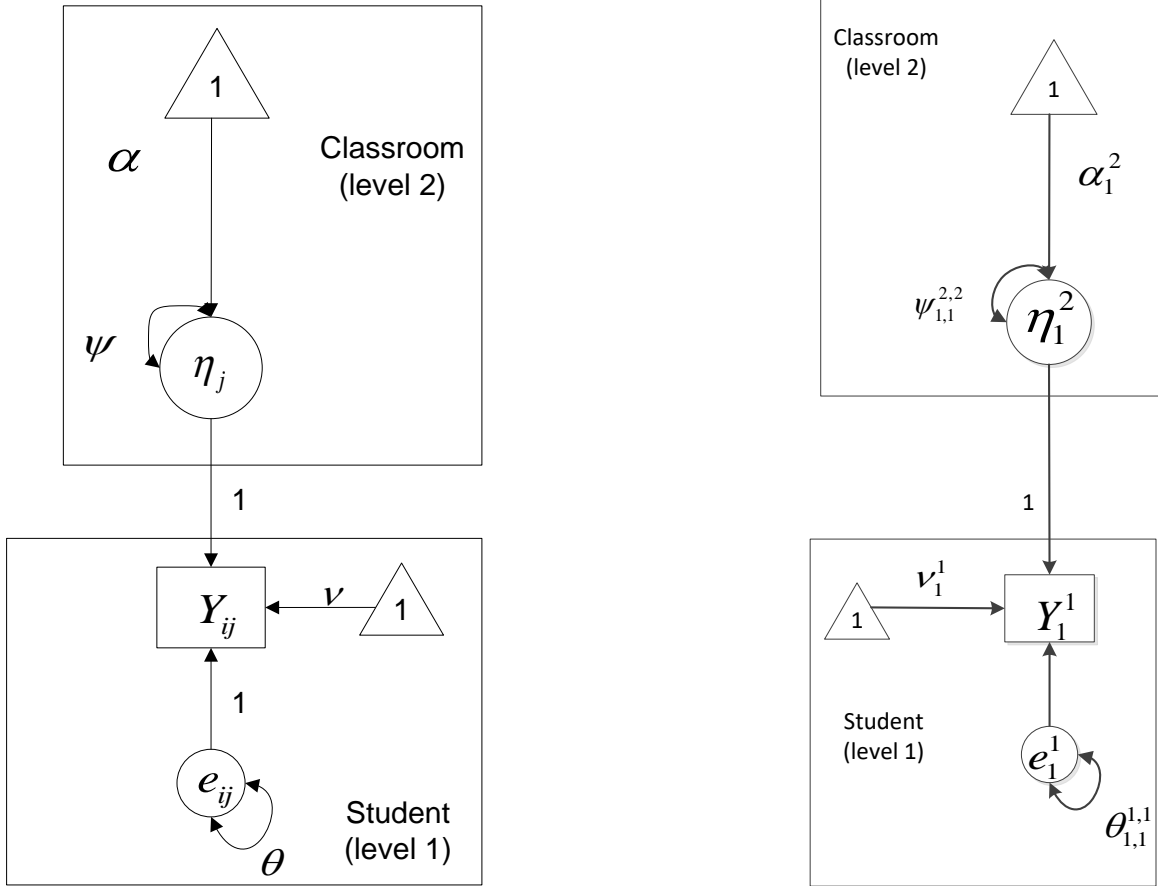


Figure 2. Graphical representations of SEM (left) and *n*SEM (right) for a two-level model.

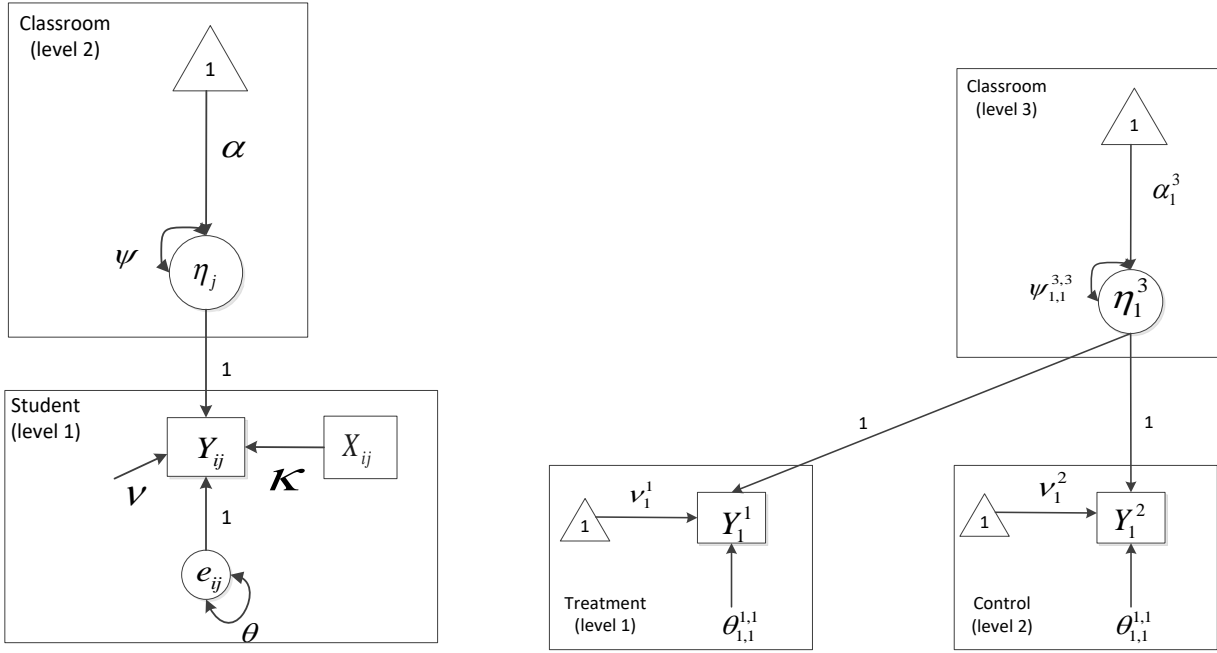


Figure 3. Conditional SEM (left) and conditional nSEM (right) measurement representations of a fully-nested randomized controlled trial.

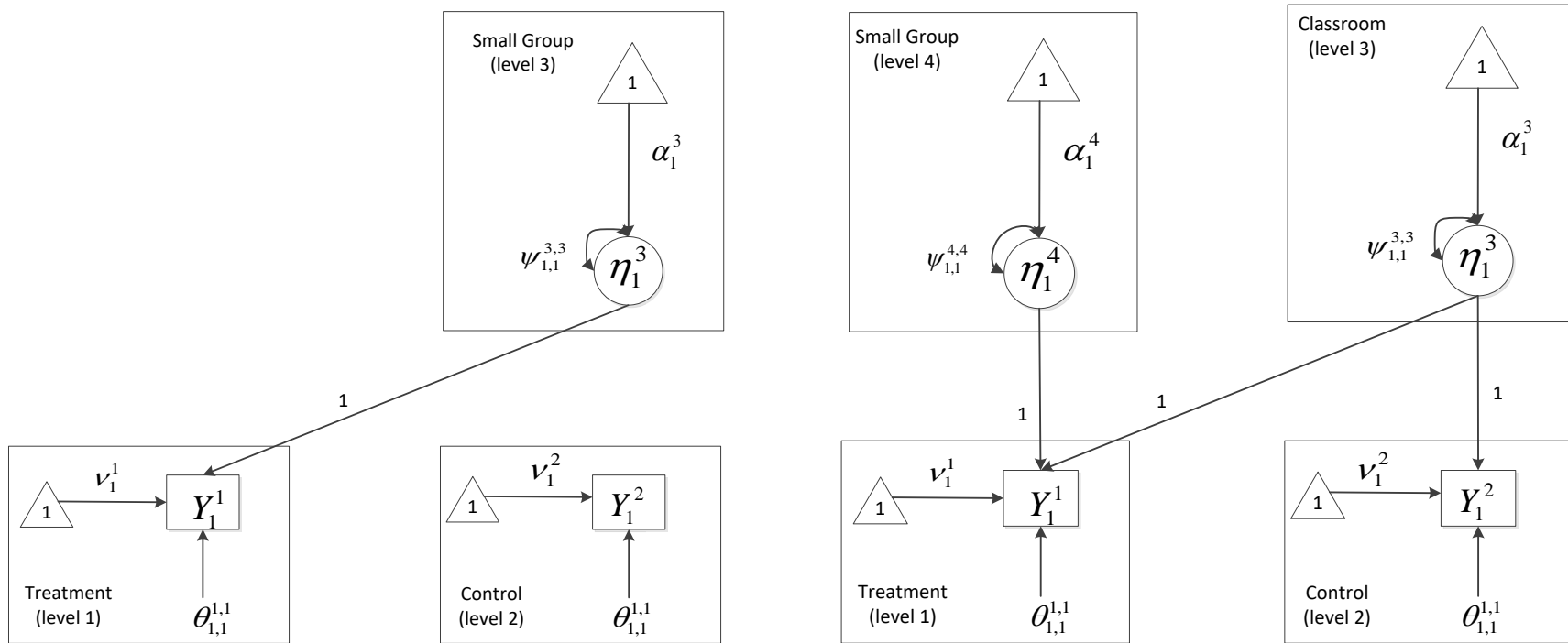


Figure 4. nSEM representations of partially nested (left) and partially nested, cross-classified (right) randomized controlled trial structures.

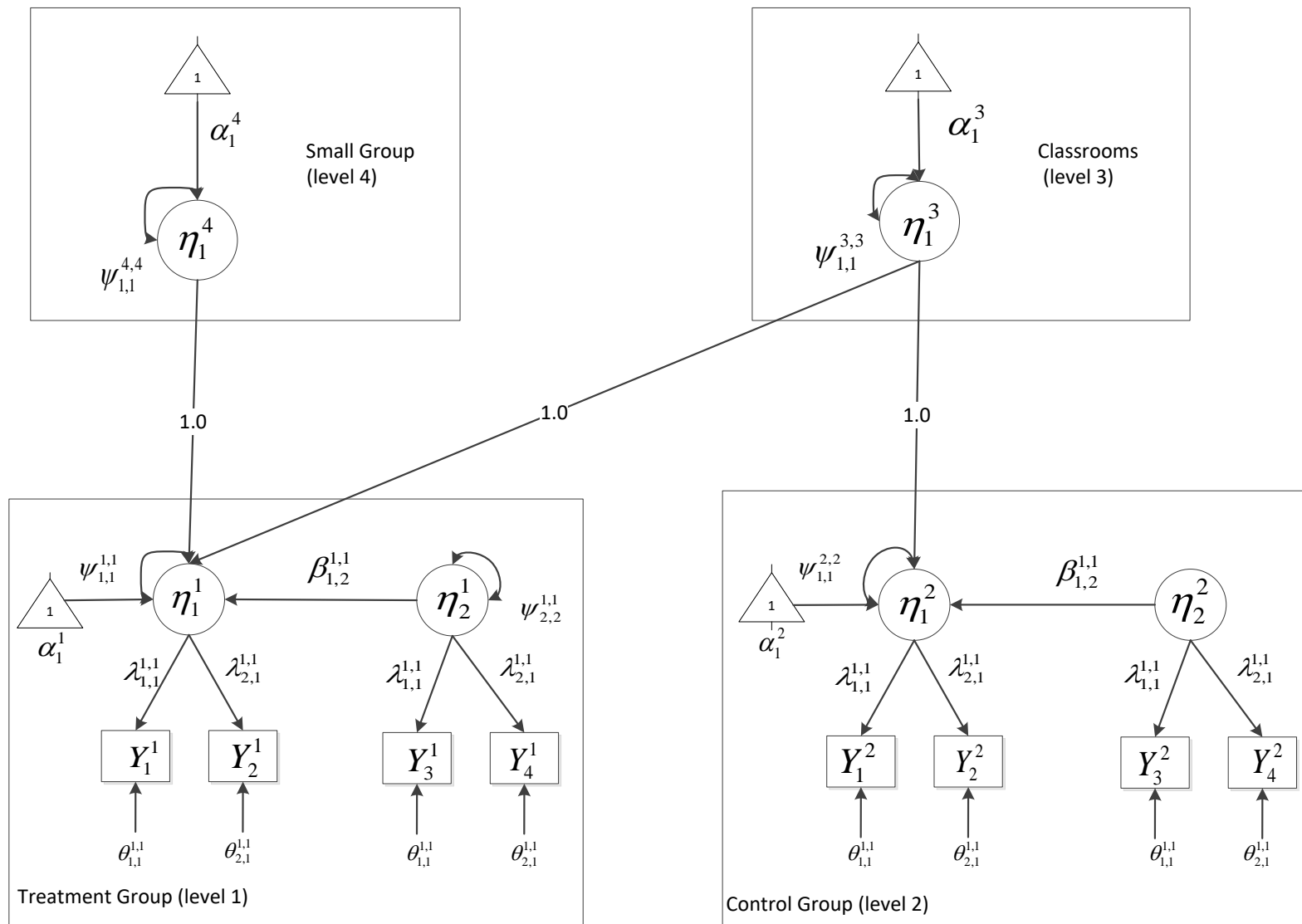


Figure 5. nSEM representations of latent partially nested, cross-classified randomized controlled trial structure.

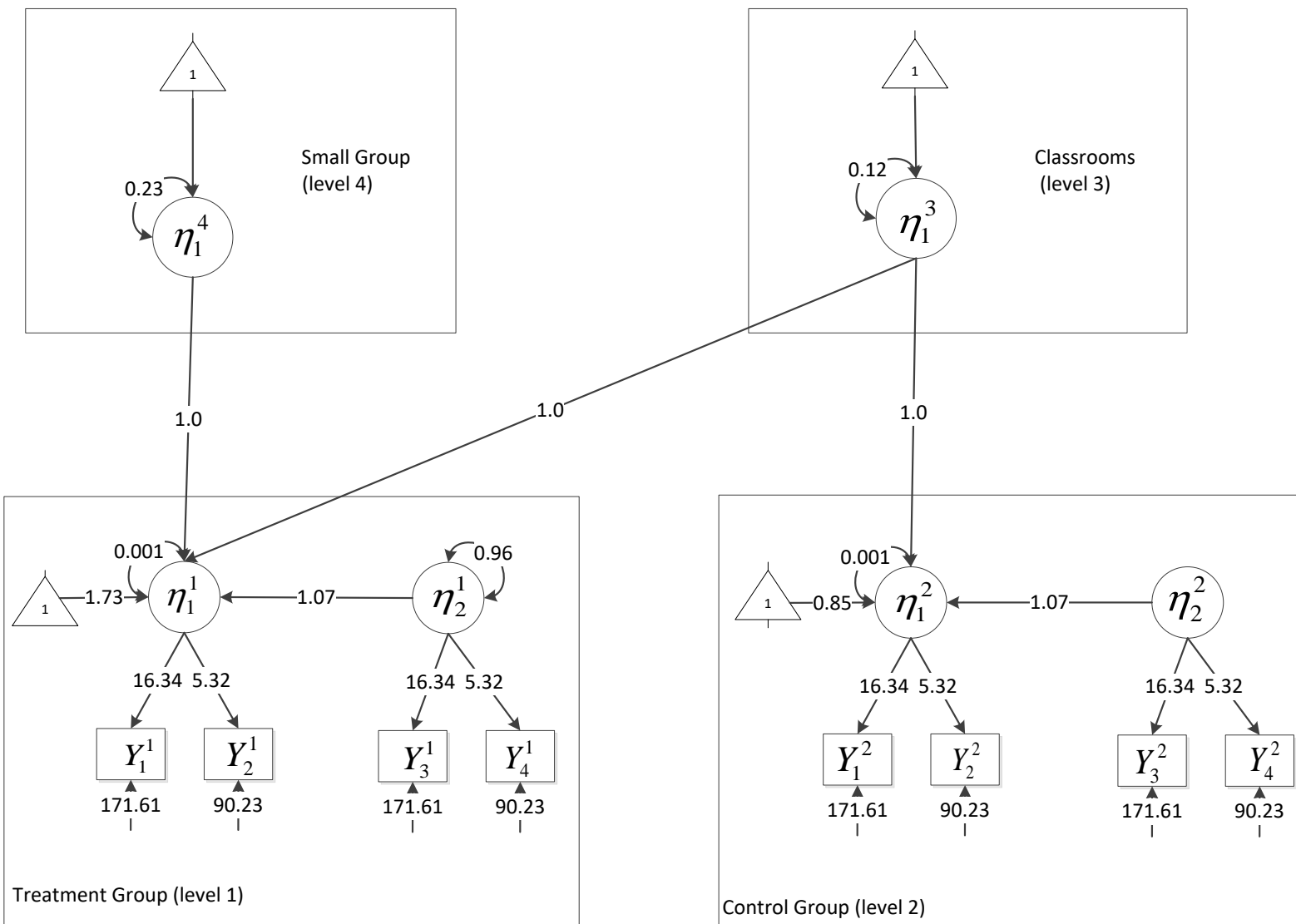


Figure 6. nSEM result for latent partially nested, cross-classified randomized controlled trial.