

People Are Variables Too: Multilevel Structural Equations Modeling

Paras D. Mehta
University of Houston

Michael C. Neale
Virginia Commonwealth University

The article uses confirmatory factor analysis (CFA) as a template to explain didactically multilevel structural equation models (ML-SEM) and to demonstrate the equivalence of general mixed-effects models and ML-SEM. An intuitively appealing graphical representation of complex ML-SEMs is introduced that succinctly describes the underlying model and its assumptions. The use of definition variables (i.e., observed variables used to fix model parameters to individual specific data values) is extended to the case of ML-SEMs for clustered data with random slopes. Empirical examples of multilevel CFA and ML-SEM with random slopes are provided along with scripts for fitting such models in SAS Proc Mixed, Mplus, and Mx. Methodological issues regarding estimation of complex ML-SEMs and the evaluation of model fit are discussed. Further potential applications of ML-SEMs are explored.

Keywords: multilevel modeling, structural equations modeling, multilevel-structural equations modeling, multilevel-confirmatory factor analysis, mixed effects models

Multilevel modeling (MLM) and structural equations modeling (SEM) have evolved from different conceptual and methodological roots. The former deals with the analysis of clustered data (e.g., students nested within classrooms) and attempts to partition observed variance into within- and between-clusters components.¹ The latter deals with modeling means and covariances among multivariate data. The two classes of models focus on different questions

and have different strengths and weaknesses. Because both clustered and multivariate data are common and of much substantive interest, it is not surprising that practitioners of each method are interested in borrowing strengths of the other (Goldstein & McDonald, 1987; Krull & MacKinnon, 2001; McDonald, 1993; Muthén, 1989; Muthén & Satorra, 1995; Raudenbush, 1995; Raudenbush & Sampson, 1999). Although considerable advances have been made in this respect, structural modeling of multilevel data is a relatively new area of methodological research (Bauer, 2003; Bentler & Liang, 2003; Curran, 2003; du Toit & du Toit, 2003; Kaplan & Elliot, 1997; Muthén, 1991, 1994, 1997; Rovine & Molenaar, 2000).

This article develops the notion of multilevel modeling as well as multilevel structural equations modeling (ML-SEM) from within the SEM framework by borrowing common and well-understood metaphors from measurement. In the broadest sense, ML-SEM combines the best of both worlds. It allows full-blown SEM models to be developed at each level of nesting for clustered data. For example, multiple individual-level indicators can be used to define a latent

Paras D. Mehta, Texas Institute for Measurement Evaluation and Statistics, University of Houston; Michael C. Neale, Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University.

This work was supported by National Institute of Child Health and Human Development Grant HD30995 (Early Interventions for Children with Reading Problems), the National Institute of Child Health and Human Development, U.S. Department of Education's Institute of Education Sciences Grant HD39521 (Oracy/Literacy Development of Spanish-Speaking Children). Michael C. Neale was supported by National Institute of Health Grants MH-65322 (Psychopathology: Models, Measurement and Classification) and DA-18673 (Psychometric and Genetic Assessments of Substance Use).

We would especially like to thank Lee Branum-Martin, Patrick W. Taylor, and Chris D. Barr for their helpful comments on earlier versions of this article.

Additional materials are on the Web at <http://dx.doi.org/10.1037/1082-989X.10.3.259.supp>

Correspondence concerning this article should be addressed to Paras D. Mehta, Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204-5022. E-mail: paras.mehta@times.uh.edu

¹ The term *nesting* and *clustering* are commonly used in the multilevel modeling literature to indicate data-nested structures such as students clustered or nested within classrooms. This notion of nesting is distinct from the notion of nested designs in an analysis of variance context in that in MLM, the sampling units at both levels are assumed to be sampled randomly. Hence, classrooms are assumed to be sampled randomly from a universe of all possible classrooms, and students are also assumed to be sampled randomly within each classroom.

construct at the individual level as well as at the cluster level, and, mediational hypotheses can be evaluated at both levels. Finally, the article demonstrates how random slopes may be incorporated into ML-SEM for clustered data.

Repeated measures data are inherently multilevel, with repeated measurements on an outcome measure nested within each individual. It is therefore not surprising that models of individual growth were initially conceptualized within the MLM framework (Raudenbush & Bryk, 2002). It was in the context of latent growth curves (LGC) that Willett and Sayer (1994) demonstrated how models of individual change could be fitted using conventional means and covariance structure analysis with time-structured data (see also Chou, Bentler, & Pentz, 1998; MacCallum, Kim, Malarkey, & Keicolt-Glaser, 1997; Meredith & Tisak, 1990). A critical assumption of the means- and covariance-based LGCs is that the time interval between measurement occasions must be equal across individuals. The availability of raw data or full-information maximum likelihood estimation (FIML) in popular SEM software partially alleviated the limitation of equal time intervals (Arbuckle, 1996; Neale, 2000b). The missing-data handling capability of FIML was exploited in order to allow unequal intervals between measurement occasions in growth curve analysis (Duncan, Duncan, & Hops, 1996; McArdle & Anderson, 1990). In this missing data approach, an outcome measured at each time point is treated as a separate variable. Hence, each individual will have data at some but not at all time points, allowing unequal intervals between measurement occasions. This approach is impractical when there is considerable heterogeneity in time of the first occasion or in the intervals between occasions of measurement. Mehta and West (2000) demonstrated the use of the individual-likelihood based SEM estimation along with the special data-handling facility for estimating the parameters of LGC models with continuous values of time. The current article extends the idea of estimating random slopes for continuous predictors to clustered data (e.g., students nested within classrooms) and multivariate outcomes.

The missing data approach has also been extended to the context of multilevel modeling of clustered data. In a set of related articles, Bauer (2003) and Curran (2003) demonstrated the applicability of SEM models to clustered data in the context of both balanced and unbalanced designs. Bauer applied two alternate approaches to missing data in longitudinal setting (Mehta & West, 2000) to the clustered data case. Both Bauer and Curran provided practical examples of how such models can be fit in conventional software. Consistent with Rovine and Molenaar (2000), Bauer demonstrated the equivalence between SEM and univariate mixed-effects model (MEM) matrices. Curran graphically presented possible extensions of the univariate random-intercepts model to the case of multiple latent variables at the individual and the cluster levels.

The primary purpose of this article is to introduce the

ML-SEMs in a didactic fashion using concepts that are well understood in the context of conventional SEM. This is accomplished by developing core ML-SEMs as variants of the restricted confirmatory factor analysis (CFA) in which factor loadings are not estimated but are fixed by design. We show how reframing such MLMs as restricted CFA models readily allows parameter estimation by individual likelihood. Finally, we show how person-specific data may be used for modeling means, as well as covariances, at an individual level. As is explained later, the ability to model person-specific covariances is central to modeling random slopes in conventional MLMs.

Several alternative representations exist for specifying and estimating MLMs of varying degrees of complexity within specific software such as HLM (Bryk, Raudenbush, Seltzer, & Congdon, 1988), MLwiN (Rasbash et al., 2000), and SAS Proc Mixed (SAS Institute, 1996). The mapping of one specification onto another, especially for multivariate outcomes, is not obvious to the uninitiated. ML-SEMs represent a considerable advance over conventional MLMs, considering the fact that the majority of applications of multilevel modeling are restricted to univariate outcomes. As such, representation of MLMs within the multivariate SEM-oriented scripting language such as Mplus (Muthén & Muthén, 2004) is even less obvious.

The current article aims to clarify the interrelationships among the different representational schemes by graphically using an extended RAM (recticular action model) representation (McArdle & Boker, 1990) of an ML-SEM.² Scalar representations of univariate and multivariate MLM map directly on to this multilevel graphical representation of a restricted CFA model. Additionally, the SEM matrices for the CFA model bear a one-to-one correspondence with the matrices of the general MEM. Finally, the notion of a random slope for a continuous predictor fits within this extended graphical representation simply as a fixed factor

² There are several alternative but highly similar graphical representations of SEM. The RAM notation is similar to other notation in that it uses rectangles and circles to represent observed and latent variables, respectively. In addition, like other representations, it uses curved double-ended arrows and directed arrows to represent covariances and variances, respectively. Because variances may be thought of as the covariance of a variable with itself, the RAM notation uses curved double-ended arrows to represent unconditional and conditional or residual variances. Finally, the RAM notation uses a directed arrow from a triangle to an observed or latent variable to represent unconditional or conditional mean. The triangle represents a fixed constant of 1.0 for each individual; hence, the directed path can be interpreted as the mean (or the intercept). An important consequence of RAM notation is that all of the model parameters are explicitly represented in a path diagram. As a result, the graphical representation can be computationally transformed into its underlying mathematical representation (see Neale, Boker, Xie, & Maes, 2004).

loading. It is hoped that the universal graphical representation will help researchers familiar with either class of models to conceptualize and fit models within the more general ML-SEM context.

ML-SEM: The Building Blocks

Three critical ideas necessary for understanding ML-SEMs are (a) the concept of a restricted CFA model, (b) the notion of individual likelihood, and (c) the related idea of modeling individual-specific data vectors. These building blocks are used to construct ML-SEMs for clustered data in a bottom-up fashion.

CFA

The pattern of means and covariances of MLMs is identical in form to that of CFA. CFA can therefore serve as a template for formulating MLMs. The relation between the two classes of models has been recognized in univariate and multivariate SEM formulations of latent growth (MacCallum et al., 1997; Willett & Sayer, 1994) as well as explicit MLM models (Rovine & Molenaar, 2000). At a practical level, the template can be easily adapted to specify and fit more complex MLMs. For example, multilevel latent variable models have been conceptualized as hierarchical factor models (Bauer, 2003; Curran, 2003).

Apart from the similarity in the data structure (i.e., nesting and unbalanced data) and the scalar representations of growth curves and multilevel regression models, there are important differences between the two classes of models. For example, unlike variables ordered in time, individuals within a cluster do not have a specific order. Such similarities and differences among different MLMs lead to a specific set of expectations regarding the structure imposed on the data as well as distributional assumptions (e.g., conditional independence and homoscedasticity) for the latent variables and residuals at each level. To the extent that the chosen factor model template adequately captures the underlying assumptions of MLMs, the use of CFA as a template is appropriate. The current article summarizes the critical implicit assumptions in SEM models for clustered data. For example, the hierarchical factor model representation of latent variable models of multivariate clustered data implies a set of falsifiable restrictions on the general-specific factor model (Gustafsson, 2002; Gustafsson & Balke, 1993).

CFA aims to explain the covariance among p observed variables (Y_p) in terms of q latent variables (η_q). All unspecified sources of variability including error and specific factors are labeled as residuals, e_{pi} :

$$Y_{pi} = \nu_p + \lambda_{pq}\eta_{qi} + e_{pi}, \quad (1)$$

where for an individual i , Y_{pi} is the score for the p th observed variable, and η_{qi} is the score for the q th latent variable, and e_{pi} is the residual or the error term for the p th

observed variable. The factor loadings (λ_{pq}) reflect the magnitude of change in the observed variable (Y_p) for unit change in the latent variable (η_q), and ν_p is the measurement intercept. The model implied mean vector is

$$\mu = \nu + \Lambda\beta, \quad (2)$$

where β is a $q \times 1$ vector of latent variable means,³ and the covariance Σ ($p \times p$) is

$$\Sigma = \Lambda\Psi\Lambda' + \Theta, \quad (3)$$

where Ψ ($q \times q$) is the latent variable covariance matrix, and Θ ($p \times p$) is a diagonal matrix containing residual variances.

Individual Likelihood Based SEM

The second building block necessary for understanding ML-SEMs is the notion of fitting SEM models to individual data vectors. The conventional means- and covariance-based SEM is limited to complete data. Cases with missing data are typically eliminated in a listwise fashion. FIML estimation was introduced in SEM as a means of handling missing data (Arbuckle, 1996; Neale, 2000b). The bulk of discussion regarding the limitations of the SEM approach to clustered data has focused on its inability to deal with unbalanced data (Bauer, 2003; Curran, 2003; Mehta & West, 2000; Willett & Sayer, 1994). This is not a limitation of the model per se, but instead reveals two related issues contributing to the inflexibility of the SEM framework: (a) the use of the sample means- and covariance-based estimation approach and (b) data handling and structuring issues that make such models difficult to formulate in a script and also make such models computationally inefficient to estimate. The availability of estimation methods based on individual likelihood makes it possible to specify and estimate fairly complex models that naturally accommodate unbalanced data structures. In this approach, likelihood is computed using individual data vectors. Assuming multivariate normality among the observed outcomes, the likelihood of a specific response vector is

$$Prob(y_i) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp(-0.5(y_i - \mu)' \Sigma^{-1} (y_i - \mu)),$$

where μ is a vector ($p \times 1$) of means of the variables, and Σ is the covariance matrix ($p \times p$); $|\Sigma|$ and Σ^{-1} are the determinant and inverse of the covariance matrix, respectively. For any given μ and Σ , twice the negative log likelihood of an individual data vector (y_i) is

³ Latent variable mean is typically represented by the Greek symbol α . In this article, we use β to represent latent variable means in order to be consistent with the fixed-effects vector in a multilevel model.

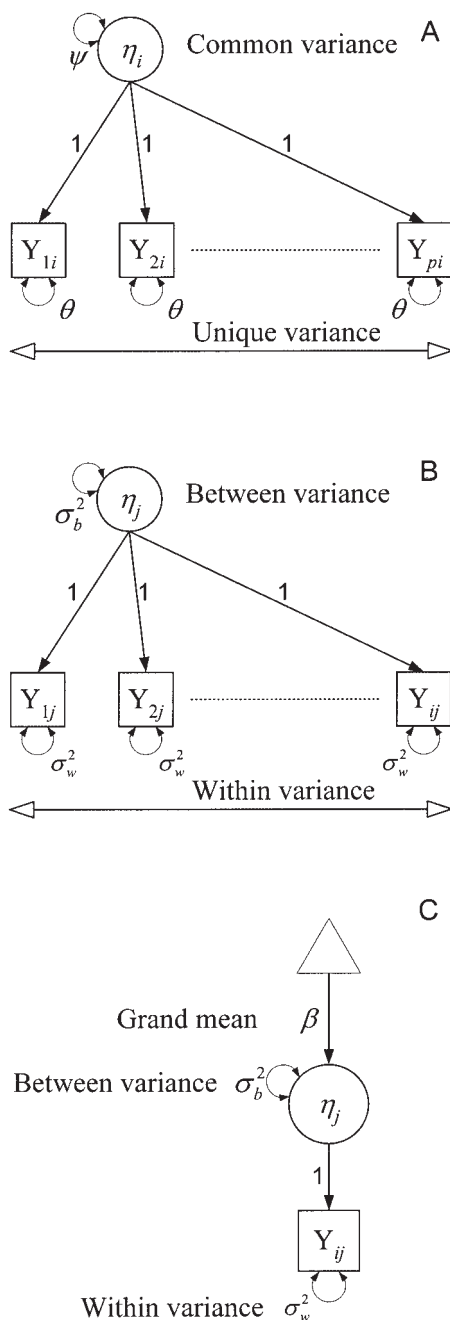


Figure 1. A: Restricted confirmatory factor analysis (CFA) model of parallel tests. Observed scores of all p variables (Y_{pi}) for individual i load on a common individual-level latent variable (η_i) with all factor loadings equal to 1.0. Dashed line indicates variables not included in the figure. Residual variances for all p observed variables are assumed to be equal (θ). Latent factor mean is not represented in the figure. B: Univariate random-intercepts model represented as a restricted CFA model. Observed scores of all individuals (Y_{ij}) in cluster j are allowed to load onto a common cluster-level latent variable (η_j) with all factor loadings equal to 1.0. The cluster-level latent variable represents the random effect for the j th cluster. Dashed line indicates scores of individuals not included in the figure. Within-cluster residual variance is assumed to be equal for all individuals. Latent factor mean (i.e., the grand

$$-2\text{Log}L_i = K_i + \log|\Sigma_i| + (\mathbf{y}_i - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i), \quad (4)$$

where $K_i = p_i \log(2\pi)$. The ML fit function for an independent sample of observations is obtained by summing the individual $-2LL$ over all the number of individuals: $-2\text{Log}L_S = \sum_{i=1}^N (-2\text{Log}L_i)$. The person-specific likelihood allows missing data without listwise deletion. Under the assumption of missing at random, conditional on the observed data, the expected individual-specific means and covariances for the variables that are missing for an individual are equal to the corresponding means and covariances for an individual with complete data. During estimation, the model implied mean and covariance matrices are computed for each unique response pattern. The variables that are missing are simply eliminated from the individual's mean and covariance matrix. In other words, the dimension of the covariance matrix varies across individuals, depending on the number of observations actually present for that individual.

Modeling Individual-Specific Covariances: Definition Variables

There is an equally important but little recognized by-product of estimation based on individual likelihood. FIML offers the possibility of fitting new types of models in which the model-implied means and covariances are different for each individual. The notion of modeling means in terms of the predictors is central to ordinary multiple regression as well as SEM; however, the notion of modeling covariance as a function of predictors is fairly novel. In SEM, multiple-group modeling is used to account for differences in covariances across discrete grouping variables. It is also plausible that the covariances among dependent variables may vary as a function of continuous-valued predictors. For example, the covariance (and variances) between latent variables stress and symptoms may be continuous functions of neuroticism:

$$\sigma_{(\text{stress}, \text{symptoms})i} = f(\text{neuroticism}_i).$$

In this case, there is no single sample covariance matrix. Instead, there are as many covariance matrices as there

mean of the dependent variable) is not represented in the figure. C: Succinct representation of univariate random-intercepts model. The figure represents the model for a single individual (i) nested in cluster (j). The within- and between-clusters variables are identified by appropriate subscripts. The observed variable includes both person (i) and group (j) subscripts, indicating that the observed scores of all individuals within a given cluster load onto the common cluster-level latent variable (η_j) with all factor loadings equal to 1.0. It is assumed that the cluster-level latent variable (η_j), and the individual-level residuals (ϵ_{ij}), are independently and identically distributed with variances σ_b^2 and σ_w^2 , respectively. This critical assumption allows us to represent the model succinctly.

are distinct values of neuroticism present in the data. Hence, it might not just be impractical but impossible to fit a separate model for each discrete value of the predictor. The individual likelihood allows person-specific means and covariances, so it is possible to compute the likelihood separately for each individual, based on their expected means and covariances, conditional on the observed values of individual predictors.

Exogenous predictors used in this fashion are called “definition variables.” Variables flagged as definition variables can be used to fix the values of any model parameter, including factor loadings and variances. Such a facility is essential for fitting models with heteroscedastic errors and interaction between latent and observed predictors (Muthén & Asparouhov, 2002, 2003; Neal, 1998, 2000a, 2000b). Other examples of the use of definition variables can be found in Eaves, Neale, and Maes (1996), Neale (2000a, 2003), and Silberg, Rutter, Neale, and Eaves (2001). Random slopes in MLM also implies that covariances are a function of some predictor. Hence, FIML is central to fitting ML-SEMs with random slopes. In the next section, we consider MLMs for a single outcome and illustrate how the random intercepts MLM has the same form as the model of parallel tests in classical test theory (CTT). Hence, the random-intercepts MLM may be fitted as a restricted CFA model.

Univariate Multilevel Models Are Really Multivariate Unilevel Models

The notions of between-clusters and within-cluster variability in MLM parallel those of common and unique sources of variability, respectively, in measurement models (Skrondal & Rabe-Hesketh, 2004). In fact, the model of parallel tests is mathematically identical to the random intercepts model in MLM. Practically, this involves switching subscripts of a univariate MLM (people and clusters) with those of a CFA model (variables and people); that is, individual scores are conceptualized as separate variables, and the unit of analysis for the SEM model is now the cluster.

Model of Parallel Tests: Common and Unique Sources of Variance

In CTT, the observed score Y_{pi} for an individual i on a test p is thought of as being composed of two components: true score and random error.

$$Y_{pi} = T_i + e_{pi}, \quad (5)$$

Note that Equation 5 is identical to Equation 1 for the general CFA model in which the intercepts and factor loadings are fixed to 0.0 and 1.0, respectively, for all p tests:

$$Y_{pi} = 0.0 + 1.0\eta_i + e_{pi},$$

where the latent variable (η_i) represents the unobserved true score. Figure 1A presents the CFA model of parallel tests. All p tests share a common origin (i.e., equal intercepts; $\nu_p = 0.0$) and have the same unit of measurement (i.e., equal loadings; $\lambda_p = 1.0$). The true score is assumed to be distributed independently and identically, $\eta \sim N(\boldsymbol{\beta}, \boldsymbol{\Psi})$, and the residual variance for all p tests is assumed to be equal ($\theta_{pp} = \theta$).⁴ In other words, all p tests are assumed to be exchangeable; that is, if we select a random subset of tests from the universe of all parallel tests, we could fit the restricted CFA model in order to estimate the variance attributable to the true score.

Because the factor loadings for all p observed variables are equal to 1.0, the variance attributable to the true-score equals the variance of the latent factor (ψ), often referred to as the “common variance.” The unique component (ε_{pi}) for a given test is independent of the residuals for the remaining tests and the true score (η_i), and its variance is commonly referred to as the “unique variance” (θ). The variance of each test is the sum of common variance (ψ) and the unique variance (θ).

$$V(Y_p) = \lambda_p^2 V(\eta_p) + V(e_p) = \psi + \theta.$$

The reliability of a test is the ratio of the true-score variance to the total variance:

$$rel(Y_p) = \frac{\text{True-score variance}}{\text{Total variance}} = \frac{\psi}{\psi + \theta}.$$

The covariance between any two tests is

$$CV(Y_p, Y_{p'}) = \lambda_p V(\eta) \lambda_{p'} = 1 * \psi * 1 = \psi.$$

In other words, the covariance between any two parallel tests equals the true-score variance. For three parallel tests, the implied covariance is

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{\Lambda} \boldsymbol{\Psi} \boldsymbol{\Lambda}' + \boldsymbol{\Theta} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [\psi] \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \\ &+ \begin{bmatrix} \theta & & \\ & \theta & \\ & & \theta \end{bmatrix} = \begin{bmatrix} \psi + \theta & \psi & \psi \\ \psi & \psi + \theta & \psi \\ \psi & \psi & \psi + \theta \end{bmatrix}. \quad (7) \end{aligned}$$

The structure of the covariance matrix remains the same regardless of the number of tests: with equal off-diagonal elements containing the true-score variance and the total variance along the diagonals. Finally, all p means are equal to the latent factor mean:

⁴ CFA allows for the possibility of specific factors in residuals, whereas the strict CTT assumes that errors are random. For the current purpose, the distinction is not critical, as we use CTT merely as a metaphor rather than as a model in a strict sense.

$$\boldsymbol{\mu} = \mathbf{v} + \boldsymbol{\Lambda}\boldsymbol{\beta} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \beta \\ \beta \\ \beta \end{bmatrix}. \quad (8)$$

Univariate MLM: Between-Clusters and Within-Cluster Variance

Consider the case in which individuals ($i = 1, 2, \dots, N$) nested within clusters ($j = 1, 2, \dots, n$) are measured on a single-outcome Y . Examples of such nested or hierarchical data structures include students within classrooms in an educational setting and employees within workgroups in an organizational setting. The observed score (Y_{ij}) for each individual i in cluster j can be thought of as a sum of (a) the respective cluster mean (η_j) and (b) an individual-specific deviation (e_{ij}):⁵

$$Y_{ij} = \eta_j + e_{ij}. \quad (9)$$

The average outcome (η_j) is also expected to vary across clusters:

$$\eta_j = \beta + u_j, \quad (10)$$

where β is the grand mean of the outcome across all clusters, and u_j is the deviation of cluster j 's mean from the grand mean. Equations 9 and 10 represent random variability in the outcome at the individual and the cluster level, respectively. The reduced form of Equation 9 can be obtained by substituting η_j from Equation 10,

$$Y_{ij} = \beta + u_j + e_{ij}, \quad (11)$$

Because the grand mean of Y_{ij} across clusters (β) is a constant added to the scores of all individuals, there are two sources of variability in an individual's score: variation that is due to clusters (u_j) and variation that is due to individuals within clusters (e_{ij}). As the two sources of variability are assumed to be independent, the variance of the observed score Y_{ij} is the sum of the between-clusters and within-cluster (i.e., individual) variance:

$$V(Y_{ij}) = V(u_j) + V(e_{ij}) = \sigma_b^2 + \sigma_w^2,$$

where σ_b^2 and σ_w^2 represent the between-clusters and within-cluster components of the observed variance. Equations 9 and 10 correspond to Level 1 and Level 2 equations of conventional MLM. The between-clusters variability in the outcome is commonly referred to as the "variance of the random intercept." The ratio of the between-clusters variance to the total variance is known as the intraclass correlation (ICC) and is an index of the magnitude of dependency among observations due to clustering.

$$\rho = \frac{V(u_j)}{V(u_j) + V(e_{ij})} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}. \quad (12)$$

Random-Intercepts SEM Model: People Are Variables Too

The parallel between a measurement model and the univariate MLM is obvious and has been noted before (Raudenbush & Sampson, 1999; Skrondal & Rabe-Hesketh, 2004).⁶ The key difference between the two models is the unit of analysis. The outcome variable in both instances involves two subscripts. In the measurement model, for each individual i , a single construct is assessed with p equivalent and exchangeable tests. In MLM, a single outcome is assessed for all individuals nested within cluster j . In both models, the observed outcome has two independent components: (a) variability at the level of the nesting unit (i.e., the individual for the measurement model and the cluster for the MLM) and (b) variability at the lowest level of the hierarchy (i.e., the exchangeable tests for the measurement level and exchangeable individuals for the MLM). The CFA model partitions the observed variance for Y_{pi} into a common (between person, ψ) and a unique component (within-person or test-specific variance, θ). Similarly, the MLM partitions the observed variance of Y_{ij} into between-clusters and within-cluster components. In the measurement model, conditional on the common latent factor, the test-specific unique components are independent of each other and are assumed to have equal variance. Similarly, in MLM, conditional on the random intercept, the individual residuals are assumed to be independently and identically distributed. Finally, the reliability of a parallel test or the proportion of observed variance attributable to the common factor is identical in form and interpretation to the ICC.

Fitting a random-intercepts MLM as a restricted CFA model of parallel tests. Given that the two models are identical, parameters of a random-intercept MLM can be estimated as a restricted CFA model. The only shift that needs to occur is the recognition that the unit of analysis is the cluster rather than the individual. Figure 1B presents the model of random intercepts as a restricted CFA model for parallel tests. The first and second subscripts in Figure 1A represent the test p and person i ; whereas, in Figure 1B, the corresponding subscripts represent individual i and cluster j , respectively. With this shift in the unit of analysis, the

⁵ Notation used for MLM is generally consistent with the mixed-effects formulation of the corresponding models. At the same time, we have tried to use notation that makes it easy to map MLM into SEM formulation.

⁶ Although, the current article deals with the case of continuous variables, the ideas presented are equally applicable for the case of latent variable models for discrete outcome including the item response theory (IRT). The relation between IRT models and corresponding MEMs for discrete outcomes has been recognized (Fox & Glas, 2001; Rabe-Hesketh, Skrondal, & Pickles, 2004; Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003; Skrondal & Rabe-Hesketh, 2004).

model-implied means and covariances for the restricted CFA model are directly applicable to the model of random intercepts.⁷ However, unlike the model of parallel tests, the number of participants is likely to vary across clusters. Hence, FIML estimation is essential for fitting a MLM as a CFA model. Representation of MLM for clustered data in Figure 1B follows directly from the representation of growth curves for repeated-measures data (Willett & Sayer, 1994). Such a graphical representation has been used for conceptualizing univariate and multivariate random-intercepts as well as random-slopes models for clustered data (Bauer, 2003; Curran, 2003). The differences between the two classes of model allow for the possibility of simplifying the representation. We now present a set of principles that allow a translation of algebraic and distributional implications of an MLM to a succinct but complete graphical representation.

Graphical representation of a univariate random-intercepts model. In conventional SEM, individuals are assumed to be sampled randomly and are thus assumed to be exchangeable. Hence, the graphical representation depicts the model for a single individual only. In Figure 1B, the latent variable represents the cluster level outcome η_j , and each observed outcome represents individual scores (Y_{ij}) for each individual i within cluster j . For this model, sampling units at both levels (i.e., clusters and individuals) are exchangeable. As a result, it is necessary to represent a single cluster and a single individual within that cluster. Figure 1C presents such a succinct graphical representation of a random-intercepts model for an individual i within cluster j . All three hypothesized parameters of the random-intercepts model are represented in Figure 1C, and the diagram accurately and completely captures the underlying mathematical model.^{8,9}

So far we have shown that the equations for a measurement model and a random-intercepts model are identical in form and interpretation. The presentation was primarily conceptual and used the hierarchical linear modeling (HLM; Bryk & Raudenbush, 1987) specification. The next section presents an MEM formulation of a multivariate MLM. MEM allows us to draw direct parallel among the matrices used in MLM and ML-SEM as well as the corresponding likelihood equations. Although the equivalence of MEM and SEM is rather trivial in the case of the univariate random-intercepts model, demonstration of a formal equivalence between MEM and CFA models in the general case allows us to represent complex MEMs as special cases of restricted CFA models.

The next section extends the equivalence of MEM and CFA models to the multivariate case. In this context, we generalize the extended RAM notation to incorporate multivariate covariance structure at the within- and between-clusters level. The resulting within- and between-clusters covariances form the basis of fitting SEM models at each level. We demonstrate how latent variable measurement models may be fitted at both levels. In addition, we describe

the set of assumptions necessary for defining random intercepts of the individual-level latent variables.

Multivariate Multilevel Models Are Also Multivariate Unilevel Models

The key insight presented so far was that SEM could be used to analyze univariate outcomes for clustered data: (a) by treating clusters as the unit of analysis and (b) by treating the outcome for each individual as a separate indicator of a cluster-level latent variable (i.e., random effect). For a single outcome, the focus was on partitioning the observed variance into between-clusters and within-cluster variance (i.e., σ_b^2 and σ_w^2). With multivariate outcomes and clustered data, we must consider the between-clusters and within-cluster covariance matrices (i.e., Σ^b and Σ^w ; see also Goldstein, 2003; Hox, 2002; Raudenbush, 1995; Raudenbush, Rowan, & Kang, 1991; Thum, 1997; Yang, Goldstein, Browne, & Woodehouse, 2001). A multivariate extension of the random intercepts model can also be conceptualized as a restricted CFA model. The SEM representation of a bivariate MLM for P_{ij} and Q_{ij} is presented next.

Bivariate MLM: Between-Clusters and Within-Cluster Covariance

The univariate random-intercepts model may be extended for two outcome variables as

$$Y_{ij}^P = \beta^P + u_j^P + e_{ij}^P \text{ and } Y_{ij}^Q = \beta^Q + u_j^Q + e_{ij}^Q \quad (13A \text{ \& } 13B)$$

where β^P and β^Q are the grand means of P and Q , respectively, u_j^P and u_j^Q are deviations of cluster j 's means of P and Q from their respective grand means, and e_{ij}^P and e_{ij}^Q are individual i 's deviation scores on the two outcomes from their respective cluster means. The between-clusters random effects (u_j^P and

⁷ Additional materials, including annotated scripts in SAS Proc Mixed, Mplus, and Mx, as well as various datasets, are on the Web at <http://dx.doi.org/10.1037/1082-989X.10.3.259.supp>

⁸ Note that when the assumption of independent clusters or exchangeable individuals is not justified, the graphical representation must be modified accordingly. Such instances may arise, for example, with data from multiple siblings in which birth order is important, or for dyadic couples data, or when the same teacher teaches multiple classrooms in school data.

⁹ Similar notion of exchangeability has been used for graphically representing general multilevel linear and nonlinear models in GLLAMM (Rabe-Hesketh, Pickles, & Skrondal, 2004; Rabe-Hesketh, Skrondal et al., 2004; Skrondal & Rabe-Hesketh, 2004) and WinBugs (Spiegelhalter, Thomas, & Best, 2000). Both of these are very general software allowing estimation of complex models including models with noncontinuous outcomes. The latter software uses Bayesian estimation and includes symbols for representing distributions of the unknown parameters.

u_j^Q) are assumed to be distributed multivariate normally with covariance (Σ^b). The individual-level residuals (e_{ij}^P and e_{ij}^Q) are also assumed to be distributed multivariate normally with covariance (Σ_i^w). Additionally, the assumption of homoscedasticity implies that the residual covariance is identical for all individuals ($\Sigma_i^w = \Sigma^w$). The variances are allowed to be different for each outcome at each level.

Although the scalar part of the previous equation is easy to understand, the covariance structure of the random effects is easily misunderstood. Because SEM was explicitly designed to model covariances, it is easy to understand the notions of between and within covariances within the SEM framework. Understanding how between-clusters and within-cluster covariance is represented in MLM and correspondingly in ML-SEM is critical for understanding and fitting complex models. To demonstrate the mapping between MLM and SEM, the graphical SEM formulation and the model matrices for the bivariate random-intercepts MLM model are presented next followed by the general MEM formulation demonstrating complete correspondence between the matrices used in the two representations.

SEM Representation of the Bivariate Random-Intercepts Model

Following the univariate model, the bivariate SEM treats the cluster as the unit of analysis. Figure 2A presents the core bivariate ML-SEM for the j th cluster. Each observed variable represents an outcome for an individual within the cluster. For N_j individuals with data on both outcomes, there would be $2 \times N_j$ observed variables representing the scores of each of the N_j individuals on the two variables. Corresponding to the scalar expression of the bivariate MLM, each observed variable is influenced by the corresponding cluster-level latent variables as well as the person-specific residual. The model for each outcome is identical to the univariate model presented earlier. The new elements are the covariances among the cluster-level latent variables and the person-level residuals.

The between-clusters model: Linking individuals within cluster to cluster-level latent variables. The latent variables (η_j^P and η_j^Q) represent the between-clusters components of the observed variables (i.e., the random intercepts). There are as many between-clusters latent variables as the number of distinct outcomes (n_y). The effect of between-clusters latent variables on the observed outcomes is the same for all individuals within a cluster. For each individual, the link between each of the two observed variables and the corresponding latent variables is an ($n_y \times n_y$) identity matrix: $\Lambda_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Because the number of individuals is likely to vary across clusters, the size of the factor-loading matrix will vary as well. For N_j individuals, the factor-loading matrix linking bivariate outcomes for all individuals to the cluster-level latent variables is obtained by vertically stacking the $n_y \times n_y$ identity

matrix for each person or by computing its Kronecker product with a unit vector of length N_j : $\Lambda_j = \mathbf{1}_{N_j} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. For a cluster with 2 individuals, the factor-loading matrix is

$$\Lambda_j = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

As in ordinary CFA, latent variables are assumed to be distributed multivariate normally, $\eta \sim N(\beta, \Psi)$. The Ψ matrix describes the covariance among the cluster-level latent variables as

$$\Psi = \Sigma^b = \begin{bmatrix} V(\eta_j^P) & \\ CV(\eta_j^Q, \eta_j^P) & V(\eta_j^Q) \end{bmatrix}$$

and is also referred to as the “between-clusters covariance.” The interpretation of the variance of the between-clusters latent variables η_j^P and η_j^Q is identical to that of the random intercept in the univariate case. The covariance between the between-clusters latent variables ($CV(\eta_j^Q, \eta_j^P)$) is the new element and represents the relation between the cluster means of Y^P and Y^Q .

The within-cluster model: residual covariance. The observed variables for each individual are assumed to have a unique, person-specific, within-cluster source of variance. These are represented by separate within-cluster latent variables¹⁰ (e_{ij}^P and e_{ij}^Q) for each individual. The covariance among individual residuals is called the “within-cluster covariance”:

$$\Sigma_i^w = \begin{bmatrix} V(e_{ij}^P) & \\ CV(e_{ij}^Q, e_{ij}^P) & V(e_{ij}^Q) \end{bmatrix}.$$

Within a given cluster, but across individuals, the residuals are assumed to be independently and identically distributed. As a result, the residual covariance matrix for the entire cluster is block-diagonal with as many independent ($\Sigma^w = \Sigma_i^w$) blocks along the diagonal as the number of individuals (N_j) within the cluster. With a cluster size of N_j , the residual covariance matrix is $\Theta_j = \mathbf{I}_{N_j} \otimes \Sigma^w$. For a cluster with 2 individuals, the residual-covariance matrix is simply

$$\Theta_j = \begin{bmatrix} \Sigma^w & \mathbf{0} \\ \mathbf{0} & \Sigma^w \end{bmatrix}.$$

¹⁰ For a multivariate multilevel model, individual-level residuals of different outcome are allowed to covary. Subsequently, the within-cluster SEM model is fitted to the within-cluster covariance among residuals. As a result, in ML-SEM, although not strictly necessary, it is conceptually convenient to treat the individual-level residuals as latent variables.

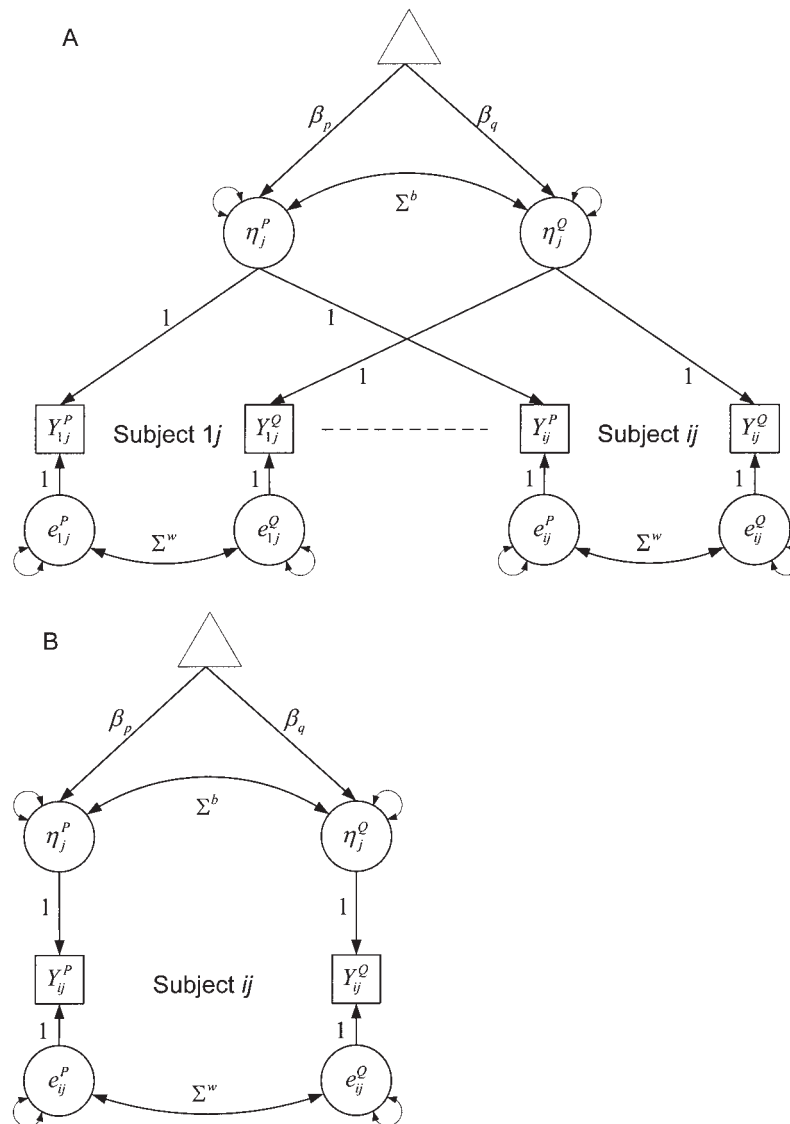


Figure 2. A: Bivariate random-intercepts model represented as a restricted confirmatory factor analysis model. Observed scores for the multivariate outcome P and Q for all individuals (i) within cluster (j) load on the corresponding between-level latent factors (η_j^p and η_j^q). Σ^b is the covariance among the between latent factors. Within-cluster residuals (e_{ij}^p and e_{ij}^q) for each individual are represented by corresponding latent variables. Σ^w is the covariance among the within-latent factors. The fixed-effects vector β (i.e., means of the cluster-level latent variables) is represented as regression paths from a triangle representing constant 1 for all individuals. Dashed line indicates the observed scores of individuals not depicted in the figure. B: Succinct representation of the bivariate random-intercepts model. The figure represents the model for a single individual (i) nested in cluster (j). The observed and latent variables as well as observed residuals include person and cluster subscripts, indicating that the level of each variable interpretation of parameters is identical to that in Figure 2A.

The SEM measurement model (Figure 2A) completely and accurately captures the scalar representation of the MLM (Equations 13A & 13B). In addition, the SEM representation clarifies the meaning of within- and between-clusters covariances that are central to multivariate MLM. The following sections demonstrate the correspondence between the SEM matrices described previously and the MEM representation of

the multivariate MLM. The most critical distinction between the SEM and MEM formulation of MLM is the fact that in SEM, the model and its likelihood are specified at the level of a cluster, whereas, in MEM, these are defined for the entire sample data vector. This difference is essentially representational as the cluster-level subset of MEM is identical to the SEM model. Identical parameters are estimated in both models

by maximizing the same likelihood. Finally, it is important to note that the size of the factor-loading and residual-covariance matrix varies across clusters, depending on the number of individuals present in the cluster. As a result, unlike conventional SEM models, these matrices have a cluster subscript. More generally, in order to correctly identify the level of a given variable, path diagrams also retain appropriate subscripts.

MEM Formulation of the Bivariate Random-Intercepts Model

The MLM equations for the two outcomes can be combined into a single equation as

$$Y_{ij}^V = D_{ij}^P \beta^P + D_{ij}^Q \beta^Q + D_{ij}^P u_{ij}^P + D_{ij}^Q u_{ij}^Q + e_{ij}^V, \quad (14)$$

where the multivariate outcomes are identified by defining two dummy variables D^P and D^Q for each individual, such that $D_{ij}^P = 1$ and $D_{ij}^Q = 0$, if the given row of observations for individual i contains Y^P and $D_{ij}^P = 0$ and $D_{ij}^Q = 1$, if the given row of observation contains Y^Q . The dummy variables D^P and D^Q serve to select appropriate fixed and random effects for each outcome variable.¹¹ The individual residual (e_{ij}^V) continues to have a variable superscript. For two outcome variables Y^P and Y^Q , the above equation for two clusters each with two individuals can be written as

$$\mathbf{Y} = \begin{bmatrix} Y_{11}^P \\ Y_{11}^Q \\ Y_{21}^P \\ Y_{21}^Q \\ Y_{12}^P \\ Y_{12}^Q \\ Y_{22}^P \\ Y_{22}^Q \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta^P \\ \beta^Q \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1^P \\ u_1^Q \\ u_2^P \\ u_2^Q \end{bmatrix} + \begin{bmatrix} e_{11}^P \\ e_{11}^Q \\ e_{21}^P \\ e_{21}^Q \\ e_{12}^P \\ e_{12}^Q \\ e_{22}^P \\ e_{22}^Q \end{bmatrix}. \quad (15)$$

Note that Equation 15 links the unobserved person and cluster-level random effects for all individuals with their respective observed scores. The SEM model in contrast represents a single generic cluster. This equation corresponds to the following general MEM formulation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (16)$$

where \mathbf{y} is a vector of the dependent variable, \mathbf{X} is the matrix of independent predictors, $\boldsymbol{\beta}$ is the unknown vector

of regression parameters or the fixed effects coefficients, \mathbf{Z} is a known design matrix used to specify the dependency structure of the observations, \mathbf{u} is the vector of unobserved random effects coefficients, and \mathbf{e} is the unobserved vector of random errors. The MEM explicitly incorporates random effects (\mathbf{u}), effectively allowing nonindependent errors. The random effects (i.e., \mathbf{u} and \mathbf{e}) are assumed to be distributed normally and independently:

$$E \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \text{ and } \text{var} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}.$$

The fixed-effects design matrix: The \mathbf{X} matrix. With two outcome variables, the fixed-effects vector has two elements (β^P and β^Q) and the corresponding \mathbf{X} matrix has two columns containing the two dummy variables (D^P and D^Q). For a given individual i with observations on both Y^P and Y^Q , the \mathbf{X}_{ij} submatrix is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and is identical to the Λ_{ij} submatrix for a given individual. Not surprisingly, the \mathbf{X}_j submatrix for cluster j is identical to the cluster-level factor-loading matrix ($\mathbf{X}_j = \Lambda_j$).

The random-effects design matrix: The \mathbf{Z} matrix. The \mathbf{Z} matrix in the bivariate case links the random effects (u_j^P and u_j^Q) for a given cluster j to appropriate rows of observations (Y^P or Y^Q) for all individuals in that cluster. In SEM, this link matrix for a generic cluster j was defined by the factor-loading matrix. In MEM, however, the link must be explicitly defined between every cluster-level random variable and every individual observed score. As a result, the number of columns in \mathbf{Z} equals the number of clusters times the number of outcome variables ($N_j * n_y$). With two outcome variables and two clusters, there are a total of four random effects. Hence, the \mathbf{Z} matrix has four columns: the first two columns link the two random effects (u_1^P and u_1^Q) for Cluster 1 with the first four rows of observations, and the last two columns link the two random effects (u_2^P and u_2^Q) for Cluster 2 with the last four rows of observations. Hence, the \mathbf{Z} matrix has as many nonzero blocks as the number of clusters. For each cluster, the \mathbf{Z} matrix has a nonzero block for observations from that cluster. Each of these blocks contains the values of the two dummy variables (D^P and D^Q) for individuals within that cluster. In the parlance of MEMs, the effects of the two dummy variables are assumed to be random at the cluster level. Note that the \mathbf{Z}_j submatrix for any cluster j has two columns for the random effect of each outcome variable and is identical to the corresponding SEM factor-loading matrix $\Lambda_j = \mathbf{Z}_j = \mathbf{X}_j$. Note that both \mathbf{X}_j and \mathbf{Z}_j submatrices are equal to the factor-loading matrix Λ_j .

¹¹ Equation 14 or Equation 15 do not appear to have an "intercept" term. This is because we now have two outcomes and therefore need an intercept term for each outcome. The grand means of the two dummy variables or their fixed effects represent outcome specific intercepts.

This is because, in SEM, the factor-loading matrix links the latent variable (η) to the observed variables for each cluster: $\mathbf{Y}_j = \Lambda_j \eta_j$. However, $\eta_j = \boldsymbol{\beta} + \mathbf{u}_j$, hence $\mathbf{Y}_j = \Lambda_j (\boldsymbol{\beta} + \mathbf{u}_j) = \Lambda_j \boldsymbol{\beta} + \Lambda_j \mathbf{u}_j$. In other words, at the cluster level, the factor-loading matrix serves as the fixed-effects design matrix (\mathbf{X}_j), as well as the random-effects design matrix (\mathbf{Z}_j).

We have shown the correspondence between SEM and MEM for all matrices except the latent factor and residual covariance (i.e., the within- and between-clusters covariances). Recall that in SEM, the model is defined for a single generic cluster j , whereas in MEM, the model is defined for all clusters and individuals. Hence, the covariance matrices for the cluster-level random effects and individual-level residuals must be defined for all n clusters and for all N individuals. These covariances are specified as the \mathbf{G} and \mathbf{R} matrices, respectively.

Covariance among cluster-level random effects: The \mathbf{G} matrix. Continuing with the bivariate example, for n clusters there are $2*n$ cluster-level random effects. Hence, the size of the covariance matrix among \mathbf{u} is $(2n \times 2n)$. For a given cluster, the between-clusters covariance is $\boldsymbol{\Sigma}^b$. This is obviously identical to the between-clusters covariance in SEM (i.e., $\boldsymbol{\Psi}$). Clusters are assumed to be sampled randomly from the population of all clusters, leading to the assumption of independently and identically distributed random effects (u_j^P and u_j^Q). As a result, across-clusters covariance among the random effects is zero ($CV(u_k, u_j) = 0$), leading to a block-diagonal \mathbf{G} matrix. The \mathbf{G} matrix has as many symmetric blocks as the number of clusters, with each block equal to the between-clusters covariance matrix ($\boldsymbol{\Sigma}^b$). With two clusters, the \mathbf{G} matrix has two symmetric blocks:

$$\mathbf{G} = \begin{pmatrix} \boldsymbol{\Sigma}^b & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^b \end{pmatrix}. \quad (17)$$

In general, with a total of n clusters, the \mathbf{G} matrix is $\mathbf{G} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}^b$, where \mathbf{I}_n is an $(n \times n)$ identity matrix. Note that the cluster-level submatrix of \mathbf{G} is identical to the SEM latent variable covariance matrix ($\boldsymbol{\Psi}$).

Covariance among individual residuals: The \mathbf{R} matrix. For a total sample size of N , there are $2*N$ individual-level residuals. Hence, the size of the residual covariance matrix is $(2N \times 2N)$. For a given individual, the within-cluster covariance (e_{ij}^P and e_{ij}^Q) is $\boldsymbol{\Sigma}^w$. Within clusters, individuals are assumed to be sampled randomly, leading to the assumption of independently and identically distributed residuals. As a result, residual covariance across individuals is zero ($CV(e_{mj}^P, e_{nj}^P) = 0$), leading to a block-diagonal \mathbf{R} matrix. The \mathbf{R} matrix has as many symmetric blocks as the total sample size (N), with each block equal to the within covariance ($\boldsymbol{\Sigma}^w$). With two clusters and two individuals within each cluster, the \mathbf{R} matrix has four symmetric blocks:

$$\mathbf{R} = \begin{pmatrix} \boldsymbol{\Sigma}^w & & & \mathbf{0} \\ & \boldsymbol{\Sigma}^w & & \\ & & \boldsymbol{\Sigma}^w & \\ \mathbf{0} & & & \boldsymbol{\Sigma}^w \end{pmatrix}.$$

In general, for a sample size of N , the \mathbf{R} matrix can be succinctly represented as $\mathbf{R} = \mathbf{I}_N \otimes \boldsymbol{\Sigma}^w$ where \mathbf{I}_N is an $(N \times N)$ identity matrix. The within-cluster submatrix of \mathbf{R} is identical to SEM residual covariance matrix ($\boldsymbol{\Theta}$).

Likelihood Equations in SEM and MEM

Given that SEM and MEM define the model at the level of clusters and the entire sample data vector, respectively, the corresponding likelihoods are also computed at the cluster and sample level, respectively. For the multivariate MEM, the mean vector is

$$\mathbf{M} = \mathbf{X}\boldsymbol{\beta},$$

and the covariance matrix among all observations is

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}.$$

The likelihood of a given vector of observations \mathbf{y} is

$$-2\text{Log}L = K + \log|\mathbf{V}| + (\mathbf{y} - \mathbf{M})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{M}),$$

where $K = N*\log(2\pi)$. For the corresponding SEM model at the cluster level, the cluster-level means and covariance are

$$\mathbf{u}_j = \Lambda_j \boldsymbol{\beta} \text{ and } \boldsymbol{\Sigma}_j = \Lambda_j \boldsymbol{\Psi} \Lambda_j' + \boldsymbol{\Theta}_j. \quad (18)$$

The likelihood of a vector of observations for cluster j , \mathbf{y}_j is

$$-2\text{Log}L_j = K_j + \log|\boldsymbol{\Sigma}_j| + (\mathbf{y}_j - \boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_j^{-1}(\mathbf{y}_j - \boldsymbol{\mu}_j),$$

where $K_j = N_j*\log(2\pi)$. Because the clusters are assumed to be sampled independently, the log likelihood for the entire sample can be obtained as the sum of cluster $-2\text{Log}L_j$ across all n clusters: $-2\text{Log}L = \sum_{j=1}^n (-2\text{Log}L_j)$. In other words, the two likelihoods are identical. Note that in practice, all mixed-effects modeling software packages recognize the block diagonal structure of the \mathbf{G} matrix and compute the likelihood using the second approach (Verbeke & Molenberghs, 2000). Hence, although the model representation stacks data for all individuals and variables into a single column, and the model specification requires the inclusion of “dummy” intercepts for accommodating the multivariate data structure, the specification and computation of the likelihood is done in a computationally efficient fashion. By corollary, the data-handling capability for specifying models with nested data structures and full-information likelihood in SEM is in and of itself inadequate for efficient estimation. Efficient algorithms similar to those used in MEM for computing the likelihood at the cluster level are necessary in ML-SEM (du Toit & du Toit, 2003).

Graphical Representation of Bivariate Random-Intercepts Model

The bivariate random-intercepts model can also be represented succinctly using the notion of exchangeability. As in the case of the univariate random-intercepts model, clusters and individuals are both considered exchangeable. Hence, it is necessary to represent the model for a single cluster (j) and a single individual (i) within that cluster. However, the two outcome variables (Y^P and Y^Q) are not exchangeable. Therefore, we need to represent the two variables and their between and within covariances separately. Figure 2B succinctly depicts a bivariate random-intercepts model. The figure shows all model parameters and is an accurate representation of a bivariate random-intercepts model. As in the univariate case, it is assumed that (a) clusters and individuals are sampled randomly and independently; (b) multiple observed repeated measures for all individuals belonging to cluster j load onto the common corresponding between-clusters variables; (c) conditional on the common between-clusters latent factors, the residuals for each outcome (e_{ij}^P and e_{ij}^Q) are independent across individuals; and (d) within-cluster covariance among residuals (Σ^w) is equal for all individuals.

Multilevel Measurement and Structural Models

Once the within and between-clusters covariances have been estimated, separate model-based restrictions may be imposed on these covariances to estimate within- and between-clusters measurement and structural parameters (see also Goldstein & Browne, 2002; Heck, 2001; Kaplan & Elliot, 1997; Snijders & Bosker, 1999). This is accomplished by defining additional within and between latent variables that impose restrictions on the core within- and between-clusters covariance matrices (Longford & Muthén, 1992; Muthén, 1991). For example, we can fit a measurement model at the between and within cluster level. This involves defining measurement model at the two levels and imposing constraints on the corresponding matrices:

$$\Sigma^b = \Lambda^b \Psi^b \Lambda^{b'} + \Theta^b \text{ and } \Sigma^w = \Lambda^w \Psi^w \Lambda^{w'} + \Theta^w. \quad (19)$$

The covariance for the entire cluster is

$$\Sigma_j = (\mathbf{1}_{N_j} \otimes \mathbf{I}_{n_y}) \Sigma^b (\mathbf{1}_{N_j} \otimes \mathbf{I}_{n_y})' + \mathbf{I}_{N_j} \otimes \Sigma^w.$$

The above equations define conventional measurement models at the within- and between-clusters levels, respectively. The measurement models may be different at the two levels. The parameters of the new measurement or structural model may be estimated by maximizing the same likelihood function. So long as the model is identified,¹² any meaningful SEM may be used for imposing restrictions on the between-clusters and within-cluster covariances. The difference in $-2\text{Log}L$ for the unconstrained model and the con-

strained model may be used to evaluate the appropriateness of the restrictions imposed by the within- and between-clusters SEMs.

The multivariate, multilevel latent variable model implied by the previous specification uses the general-specific factor model as a template as opposed to the hierarchical factor model (Gustafsson, 2002; Gustafsson & Balke, 1993). The cluster-level specification represents the general part of the model, whereas the individual-level model corresponds to the specific part of the model. From this perspective, the base model (i.e., the multivariate MEM) also uses the general-specific factor model template. Bauer (2003) and Curran (2003) used a conceptually appealing but restrictive alternative, the hierarchical factor model as a template for conceptualizing multilevel latent variable models. However, the hierarchical model specification implies a specific set of constraints on the less restrictive general-specific model. The first restriction is the restriction imposed by the within- and between-clusters factor models on the within- and between-clusters covariances (Equation 19). The second set of restrictions is related to (a) the assumption of invariant factor loadings across levels and (b) the assumption of zero variances of observed indicators at the cluster level. These restrictions are illustrated next in the context of an empirical example. It should be noted that the notion of hierarchical factor models can also be used as the measurement model for the clustered data (see Harnqvist, Gustafsson, Muthén, & Nelson, 1994).

Multilevel CFA: Empirical Example

We illustrate estimation of within- and between-clusters covariances as well as fitting a measurement model at the two levels with an example from reading research. The data come from a larger longitudinal study of early reading development involving 1,052 third graders nested within 115 classrooms (Mehta, Foorman, Branum-Martin, & Taylor, 2005). The sample was primarily African American (94.58%) with Hispanics being the second largest group (4.66%). These data come from end-of-the-year assessment of literacy skills and include the following broad measures of literacy: phonemic awareness (PA), word reading (WR), spelling (S), vocabulary (V), and writing (WT). The details of the psychometric properties of these measures can be found elsewhere (Foorman et al., 2003). We consider the possibility that these diverse measures of English literacy are indicators of a single factor, both at the individual and at the classroom level. Initial univariate and multivariate MEMs were fitted using SAS Proc Mixed, and the ML-CFA

¹² The multivariate MEM is identified so long as there are enough clusters to meaningfully estimate covariances at the between-clusters level and enough individuals to estimate within-cluster covariance. Conventional rules of identification apply for fitting SEM models at each level.

model was fitted using Mplus (Version 3; Muthén & Muthén, 2004).

Univariate Random-Intercepts Model

As a first step, a univariate random-intercepts model was fitted to the data, with classroom as the Level 2 nesting unit (see Table 1). The between-classroom variability was significant for each of the five measures with ICCs (proportion of variance attributable to between-clusters variability) ranging from 0.10 to 0.24. The means of each measure are also reported in Table 1. Grand means are not particularly interesting in a single-group model and are ignored in subsequent analyses. It is obvious that univariate analyses do not provide any information regarding covariance among outcomes at the within- and between-clusters levels.

Multivariate Random-Intercepts Model

A multivariate random-intercepts model (see Figure 3) was fitted in Proc Mixed to obtain between-clusters (Σ^b) and within-cluster (Σ^w) covariance among the outcome measures. Table 2 presents these between-clusters and within-cluster covariances (below the diagonal) along with the corresponding correlations (above the diagonal). Note that the estimated within- and between-clusters variances (diagonal elements) were almost identical to those reported in Table 1 for the univariate random-intercepts model. The correlations among the outcome variables are different at the two levels (see Figure 3). Eight out of 10 correlations at the between-clusters level are larger than the corresponding within-cluster correlations. The patterns of within- and between-clusters correlations are consistent with a unidimensional model at each level. A single-factor model asserts that a single latent factor explains the between-clusters (Σ^b) and within-cluster (Σ^w) covariances; that is, conditional on the latent factors at each level, the within- and between-clusters residuals are independent.

Multilevel CFA

Table 3 presents the unstandardized parameter estimates for the ML-CFA fitted in Mplus 3.11. All five observed variables were allowed to load onto a single literacy factor

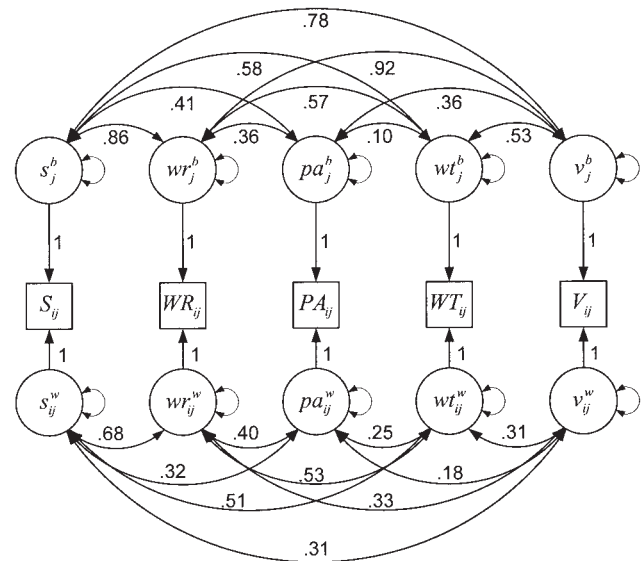


Figure 3. Multivariate random-intercepts model of literacy. The observed variables (uppercase) include person and cluster subscripts. Individual-level residuals (e_{ij}) for each outcome are represented as latent variables. Superscripts b and w indicate between-clusters and within-cluster latent variables (lower case). Between-level latent variables have a cluster subscript, whereas the within-level latent variables have both individual and cluster subscripts. The five outcome variables are spelling (S), word reading (WR), phonemic awareness (PA), writing (WT), and Vocabulary (V). The grand means of the between-level latent variables are not included for clarity. Numbers represent correlations among latent variables.

at both the within- and between-clusters level. The latent factor scales at each level were identified by fixing the WR factor loading to 1.0. The remaining factor loadings, latent factor variances, and residual variances were freely estimated at both levels. Measurement intercepts were estimated for all five outcome variables, and the mean of the between-clusters literacy factor was fixed to 0.0. Because no restrictions were imposed on the mean structure, the estimated intercepts were almost identical to the corresponding grand means in the univariate MEM. Standardized parameter estimates for the ML-CFA are presented in Figure 4.

Mplus reports within- and between-clusters estimated

Table 1
Univariate Random-Intercepts Mixed-Effects Models: Fixed and Random Effects

Outcome	Grand mean	SE	Between-clusters variance	SE	Within-cluster variance	SE	ICC
Spelling	91.38	0.65	29.23	6.39	158.99	7.41	0.15
Word reading	2.27	0.04	0.08	0.02	0.70	0.03	0.10
Phonemic awareness	0.93	0.03	0.05	0.01	0.21	0.01	0.18
Writing	-0.08	0.05 ^a	0.18	0.03	0.55	0.03	0.24
Vocabulary	78.67	0.70	28.87	7.47	229.24	10.79	0.11

Note. ICC = intraclass correlation.

^a Not significant.

Table 2

Multivariate Random-Intercepts Mixed-Effects Model: Fixed and Random Effects

Outcome	Spelling	Word reading	Phonemic awareness	Writing	Vocabulary
Grand mean	91.29 (0.65)	2.31 (0.04)	0.95 (0.03)	-0.10 (0.05)	78.57 (0.70)
Between-clusters covariance/correlation (\mathbf{G}_j) ^{a,b}					
Spelling	29.26 (6.41)	0.86	0.41	0.58	0.78
Word reading	1.49 (0.36)	0.10 (0.02)	0.36	0.57	0.92
Phonemic awareness	0.51 (0.20)	0.03 (0.01)	0.05 (0.01)	0.10	0.36
Writing	1.31 (0.37)	0.08 (0.02)	0.01 (0.01)	0.18 (0.03)	0.53
Vocabulary	22.82 (5.71)	1.60 (0.36)	0.45 (0.21)	1.21 (0.38)	29.33 (7.47)
Within-cluster covariance/correlation (\mathbf{R}_{ij}) ^{a,b}					
Spelling	158.42 (7.36)	0.68	0.32	0.51	0.31
Word reading	7.28 (0.43)	0.72 (0.03)	0.40	0.53	0.33
Phonemic awareness	1.85 (0.21)	0.16 (0.01)	0.21 (0.01)	0.25	0.18
Writing	4.83 (0.35)	0.33 (0.02)	0.09 (0.01)	0.55 (0.03)	0.31
Vocabulary	59.74 (6.63)	4.27 (0.46)	1.23 (0.25)	3.47 (0.40)	228.69 (10.73)

Note. Numbers in parentheses represent the standard error. $-2LL$ for the model is 21,507.200

^a Covariances are presented below the diagonal, and correlations are above the diagonal.

^b Diagonals elements are the variances.

sample statistics. These estimates are obtained by fitting an unrestricted model (referred to as the *H1 model*). The H1 model is in fact identical to the multivariate MEM presented previously (see Figure 3). Explicitly fitting MEM clarifies the source and the meaning of the sample statistics reported by Mplus. Mplus parameter estimates and the fit statistics for the H1 model were identical to those reported by Proc Mixed for the corresponding multivariate random-intercepts model (see Table 2).

The Mplus log likelihood for the H1 model was -10,753.615. The corresponding Proc Mixed $-2LL$ for the multivariate MEM was 21,507.2 ($= -2 \times -10,753.615$). The Mplus log likelihood for the restricted (H0) model was 10,761.576. Given certain regularity conditions, the difference in $-2LL$ between the H0 and H1 model is distributed as a chi-square with degrees of freedom equal to the difference in the number of parameters estimated by the two models. In this case, the overall fit statistic was $\chi^2_{10} = 15.92$ ($p = .10$), suggesting that the restriction imposed by the ML-CFA model on the within- and between-clusters covariances did not result in a worse fitting model. Standardized root mean squared residual for the between and within models were 0.043 and 0.019, respectively, suggesting that the multilevel model did an adequate job in reproducing covariances at both levels.

The estimated variance of the latent literacy factor at the within- and between-clusters level was 0.52 and 0.10, respectively. However, in the absence of a common scale, the magnitude of these variances is not directly comparable. The issue of establishing common across-level scale is discussed in the next section. The proportion of variance explained by the latent literacy factor at the within level ranged from .16 to .72. The corresponding proportions at the

between level ranged from .15 to .96. The relatively low proportion of variance explained at the within level suggests that after controlling for classroom variability in literacy, there is considerable heterogeneity in the within-individual variability in literacy. At the between level, PA and WT had the lowest R^2 . It is generally acknowledged that PA may be a less important indicator of literacy beyond second grade. This could explain the low R^2 for PA at both within- and between-clusters level. On the other hand, low R^2 for WT at

Table 3

Multilevel Confirmatory Factor Analysis: Unstandardized Parameter Estimates

Measure	Factor loading	Residual variance	R^2
Within cluster ^a			
Spelling	13.91 (0.65)	57.32 (4.55)	0.64
Word reading	1.00	0.20 (0.02)	0.72
Phonemic awareness	0.28 (0.02)	0.17 (0.01)	0.19
Writing	0.65 (0.04)	0.33 (0.02)	0.40
Vocabulary	8.47 (0.75)	191.20 (9.33)	0.16
Between clusters ^b			
Spelling	15.16 (1.91)	6.28 (2.63)	0.78
Word reading	1.00	0.00 (0.01) ^c	0.96
Phonemic awareness	0.28 (0.10)	0.04 (0.01)	0.15
Writing	0.81 (0.17)	0.11 (0.02)	0.36
Vocabulary	15.72 (2.50)	5.02 (4.40)	0.83

Note. Numbers in parentheses represent the standard error. $-2LL$ for the model was 21,523.152.

^a Variance of the literacy factor at the within level was 0.52 (0.04). ^b Variance of the literacy factor at the between level was 0.10 (0.02). ^c Residual variance for word reading at the between level was practically estimated at zero.

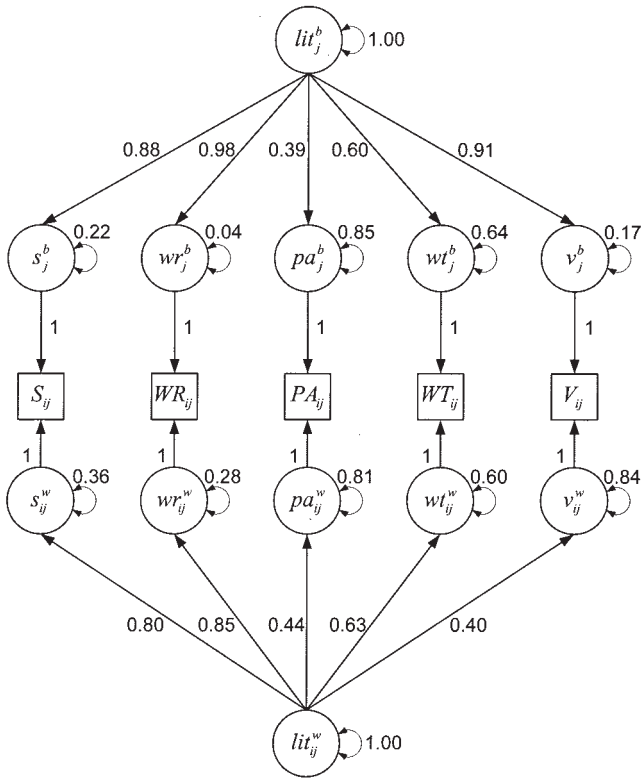


Figure 4. Multilevel confirmatory factor analytic model of literacy: completely standardized solution. The between- and within-level latent variables are allowed to load onto the corresponding between- and within-level literacy factors (lit). The five outcome variables are spelling (S), word reading (WR), phonemic awareness (PA), writing (WT), and Vocabulary (V). Note that the factor model includes residual variances at each level. The measurement intercepts of the between-level residuals are not included for clarity.

the between level may indicate that there may be some other factor influencing classroom level variability in WT.

Random Intercepts for a Latent Variable

So far we have described latent literacy factor at both the within- and between-clusters level. This raises two related questions: (a) If there can be random intercepts (i.e., variability in the cluster means for individual-level observed variables), can we have a random intercept for the individual-level literacy factor? (b) Alternatively, does the cluster-level literacy factor mean the same thing as the literacy factor at the individual level? Technically these are questions of across-level invariance of the measurement model.

Consider a two-level measurement model with p observed variables and q latent variables at each level. Each of the p observed variables may be expressed in terms of within- and between-clusters deviations:

$$Y_{ij}^p = \beta^p + y_j^{p(b)} + y_{ij}^{p(w)}.$$

These within- and between-clusters deviations serve as indicators of q th within and between latent variables:

$$y_{ij}^{p(w)} = \lambda_{pq}^{(w)} \eta_{ij}^{Q(w)} + e_{ij}^{p(w)} \text{ and } y_j^{p(b)} = \lambda_{pq}^{(b)} \eta_j^{Q(b)} + \xi_j^{p(b)}.$$

If the factor loadings are invariant across levels ($\lambda_{pq}^{(b)} = \lambda_{pq}^{(w)} = \lambda_{pq}$), then the above expression may be simplified as

$$\begin{aligned} Y_{ij}^p &= \beta^p + \lambda_{pq} \eta_j^{Q(b)} + \xi_j^{p(b)} + \lambda_{pq} \eta_{ij}^{Q(w)} + e_{ij}^{p(w)} = \beta^p \\ &+ \lambda_{pq} (\eta_j^{Q(b)} + \eta_{ij}^{Q(w)}) + \xi_j^{p(b)} + e_{ij}^{p(w)} \\ &= \beta^p + \lambda_{pq} \eta_{ij}^Q + \xi_j^{p(b)} + e_{ij}^{p(w)} \end{aligned}$$

where η_{ij}^Q is an individual-level latent variable that is itself composed of within- and between-clusters deviations: $\eta_{ij}^Q = \eta_j^{Q(b)} + \eta_{ij}^{Q(w)}$. In other words, we now have an individual-level latent variable with a random intercept at the between-clusters level. Conceptually, invariance of across-level factor loading equates the scales of the latent common factor across levels, thus making latent factor variances to be directly comparable. Also note that the observed indicator continues to have variability at the between level ($\xi_j^{p(b)}$). This is the residual for the second-level factor model.

For the literacy data, the hypothesized model with invariant factor loadings is presented in Figure 5. The literacy factor at the individual level now has a cluster-level random intercept. In addition, the observed variables continue to have cluster-level random intercepts. These are actually, the residual variances for the between-level random intercepts. As such, the random intercepts for Level 2 observed residuals are independent of the between latent literacy factor. The Mplus log likelihood for the model with invariant across-level factor loadings was $-10,753.615$. The difference in $-2LL$ between the model with and without invariant factor loadings was $\chi_4^2 = 8.618$ ($p = .071$), indicating that the hypothesis of invariant factor loadings cannot be rejected. The parameter estimates of the model with invariant factor loadings are presented in Table 4 and are generally very similar to the unconstrained model. Invariant factor loadings make common-factor variances to be directly comparable across levels. The variance of the random intercept for the literacy factor was 0.111, and the corresponding residual within-cluster variance was 0.507. The proportion of variance in the individual-level literacy factor explained by its random intercept or ICC for the latent literacy factor was .18.

The model presented in Figure 4 uses the general-specific factor model as a template, with the between-level model representing the general factor, and factor models for each individual representing the specific factors. In contrast, model in Figure 5 is analogous to the hierarchical factor model, in which latent variables at the individual-level (i.e., the first-order factors) define the latent factor at the higher level. However, note that individual observed indicators

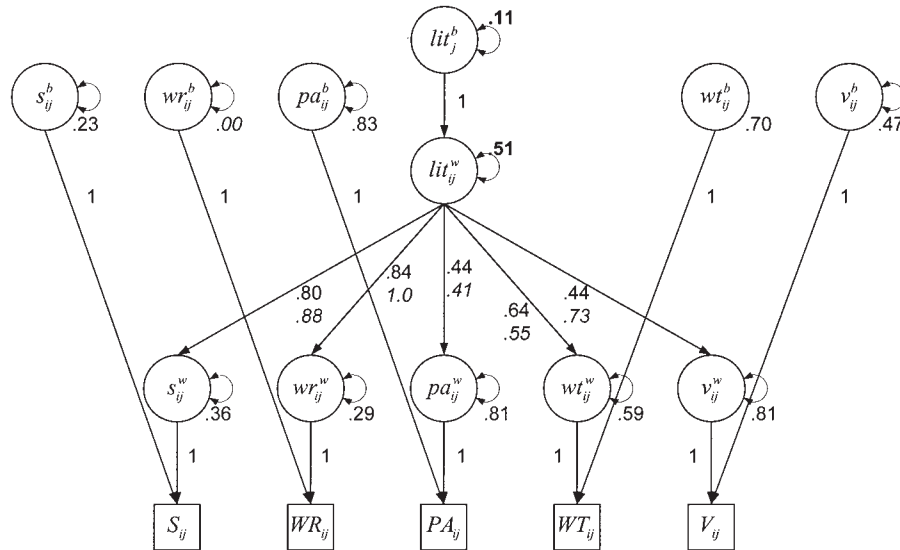


Figure 5. Random intercept for a latent variable: multilevel confirmatory factor analytic model with invariant across-level factor loadings. The factor loadings and residual variances at each level are standardized estimates. Numbers in italics are the standardized between-factor loadings. The variance of the literacy factor (*lit*; in bold) at each level is unstandardized. The five outcome variables are spelling (S), word reading (WR), phonemic awareness (PA), writing (WT), and Vocabulary (V). Note that the unstandardized factor loadings are invariant as shown in Table 4.

Table 4
ML-CFA With Invariant Across-Level Factor Loadings:
Unstandardized Parameter Estimates

Measure	Factor loading	Residual variance/standard deviation	R ²
Within cluster ^a			
Spelling	14.16 (0.58)	56.74 (4.37)	0.64
Word reading	1.00	0.20 (0.02)	0.71
Phonemic awareness	0.28 (0.02)	0.17 (0.01)	0.19
Writing	0.67 (0.04)	0.33 (0.02)	0.41
Vocabulary	9.49 (0.69)	190.43 (9.35)	0.19
Between clusters ^{b,c}			
Spelling	14.16 (0.58)	2.58 (0.45)	0.77
Word reading	1.00	0.00 (0.17) ^c	1.00
Phonemic awareness	0.28 (0.02)	0.21 (0.02)	0.17
Writing	0.67 (0.04)	0.34 (0.03)	0.30
Vocabulary	9.49 (0.69)	2.977 (0.711)	0.53

Note. Numbers in parentheses represent the standard error. $-2LL$ for the model was 21,531.800.

^a Variance of the literacy factor at the within level was 0.51 (0.04). ^b Variance of the literacy factor at the between level was 0.11 (0.02). ^c At the between-level, residual standard deviations were estimated to avoid negative residuals. ^d Residual variance for word reading at the between level was practically estimated at zero. ML = multilevel; CFA = confirmatory factor analysis.

continue to retain variability at the higher level ($\xi_j^{P(b)}$). In summary, invariance of across-level factor loadings in the general-specific model and zero variability for the observed indicators at the second level are necessary prerequisites for the use of the parsimonious hierarchical factor model (Bauer, 2003; Curran, 2003).

Random Slopes in SEM and MLM

The notion of random slopes is novel in general SEM for clustered data. Hence, we present this idea conceptually within its native MLM framework and demonstrate how the random slope may be represented as a latent variable within the familiar CFA model. The random-intercepts model can be extended to include the random effect of an individual predictor (X_{ij}):

$$Y_{ij} = \eta_{1j}1 + \eta_{2j}X_{ij} + e_{ij}, \quad (20)$$

where η_{2j} is the effect of X on Y for the j th cluster, and η_{1j} is the intercept (i.e., the level of Y_{ij} when $X_{ij} = 0$). The presence of subscript j in η_{2j} indicates that the regression coefficient is assumed to vary across clusters, hence the term random slopes. The random regression coefficients can be expressed in deviation form as

$$\eta_{1j} = \beta_1 + u_{1j} \text{ and } \eta_{2j} = \beta_2 + u_{2j}, \quad (21A \text{ \& } 21B)$$

where β_1 and β_2 are the mean intercepts and slopes, respectively, and u_{1j} and u_{2j} are the corresponding cluster deviations. Note that the intercept is defined at $X_{ij} = 0$. If X_{ij} is centered at the grand mean, then β_1 and u_{1j} are defined at the grand mean of X_{ij} .

Formulating Random Slope as an SEM Latent Variable: The Notion of a Definition Variable

For the random-intercept model, the observed variables (Y_{ij}) were the scores for an individual i nested within classroom j . The latent variable (η_{1j}) represented the random intercept. Because individuals within classrooms are interchangeable, the factor loading ($\lambda_{i,1}$) for each individual was fixed to 1.0. Similarly, intercluster variability in the effect of X on Y (η_{2j}) may be represented as a random variable. To do so, the factor loadings for η_{2j} must be fixed to the observed value of X_{ij} for each individual ($\lambda_{i,2} = X_{ij}$), which may in fact be different for every individual in the sample. Until very recently, most popular SEM software did not support any mechanism for fixing model parameters to an individual's data values. Currently, Mx (Neale et al., 2004) and Mplus¹³ (Muthén & Muthén, 2004) offer such a capability, and the variable that is used for fixing model parameters is called a "definition variable."

Graphical representation of a random-slopes model. Figure 6A presents a CFA formulation of the random-slopes model. X_{ij} is used as a definition variable in which individual-specific values of X_{ij} are used to fix model parameters (i.e., the slope factor loadings of Y_{ij} on η_{2j}). Definition variables are represented by placing the variable (X_{ij}) inside a diamond symbol (\diamond). As in the case of a random-intercepts model, because individuals are assumed to be exchangeable, it is necessary to represent the model for a single individual only. Figure 6B illustrates a CFA formulation of the random-slopes model using the succinct graphical representation introduced earlier. Two extensions to the RAM notation (McArdle & Boker, 1990) are necessary for representing a random-slopes model: (a) the inclusion of individual and cluster subscripts to incorporate multilevel information and (b) the use of diamonds to represent definition variables.

An alternate graphical representation of a random-slopes model. Random slopes are graphically represented in a somewhat different fashion in Mplus (Figure 6C). This approach does not use individual and cluster subscripts to indicate the level of each variable, instead the within and between parts of the model are represented by using explicit labels and by a line separating the two models. The outcome variable is assumed to have variability at the cluster level. The fact that the effect of X_{ij} on Y_{ij} is random at the cluster level is represented by a dot on the line representing the regression effect of X on Y (in the within part of the diagram). Within the script, the corresponding random re-

gression is specified as: $S \mid Y$ on X . The label of the random effect so defined (s) is placed next to the dot to indicate the name of the random variable at the cluster level. The random intercept at the between level is referred to by the variable label of the dependent variable. The between part of the diagram has two latent variables y (intercept) and s (slope) for the two random effects.

We prefer the extended RAM notation for representing multilevel SEM models for several reasons. First, the path diagram is a direct, complete, and mathematically accurate representation of the underlying model, and may be used for representing random slopes of any type at both levels (individual and cluster), including random slopes in the context of longitudinal data with continuously varying values of time and possibly time-varying covariates. Second, conventional rules of a path diagram may be used without any modifications to derive the implied means and covariances. Third, the extended RAM notation reformulates the notion of random slopes into the familiar concept of a restricted CFA model, yet the meaning of the diagram is readily interpretable in terms of both the multilevel random-slopes model and the restricted CFA model. For example, both intercept and slope factors are defined as conventional latent variables, with factor loadings fixed to 1.0 and the person-specific X_{ij} , respectively. Fourth, the diagram makes an explicit distinction between a modeled dependent variable (Y_{ij}) and an exogenous observed variable X_{ij} used as a definition variable.

Correspondence Between the SEM and MEM Formulation of Random Slopes

It is clear from Figures 6A and 6B that the random-slopes model can be represented as a cluster-level restricted CFA model in which the factor loadings for the slope are fixed to individual specific values of the predictors X_{ij} . For cluster j with three individuals, the factor-loading matrix is

$$\Lambda_j = \begin{bmatrix} 1 & X_{1j} \\ 1 & X_{2j} \\ 1 & X_{3j} \end{bmatrix}.$$

In general, for a cluster with N_j individuals, the factor-loading matrix is $\Lambda_j = (\mathbf{1}_{N_j} | \mathbf{X}_j)$ where $\mathbf{1}_{N_j}$ is a unit vector, \mathbf{X}_j is the predictor vector for all individuals in cluster j , and $|$ represents horizontal concatenation of the two column vectors. The two latent factors (η_{1j} and η_{2j}) represent the random intercept and random slope, respectively. The latent factor means represent the fixed effects:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix},$$

¹³ Mplus does not use the term definition variable, although such a functionality is implicitly available to a certain extent.

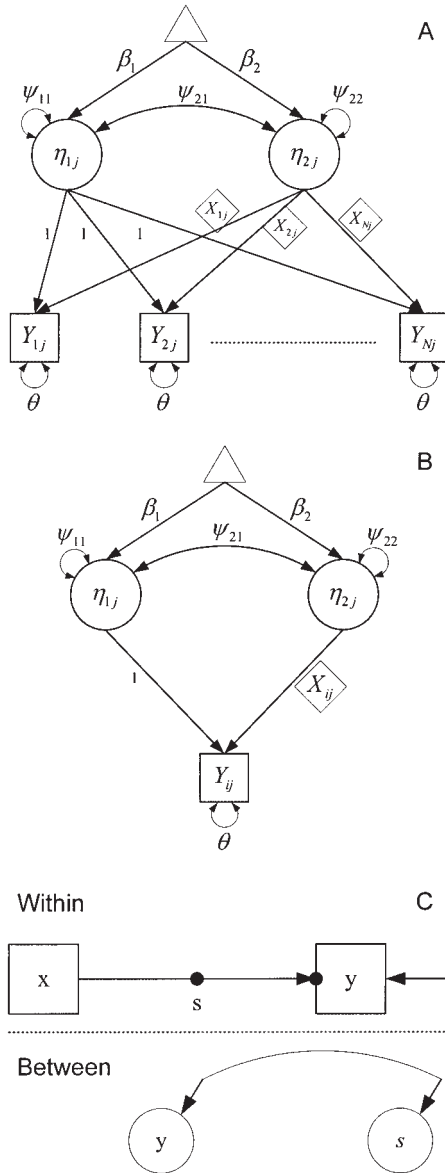


Figure 6. A: Univariate random-slopes model represented as a restricted CFA model. Observed scores of all individuals (Y_{ij}) in cluster j are allowed to load onto two latent factors: intercept (η_{1j}) and slope (η_{2j}). The factor loadings for all observed variables on the intercept factor are fixed to 1.0. The factor loading for each individual on the slope factor is fixed to individual's observed score on X_{ij} (definition variables). Definition variables are represented by a diamond symbol. Dashed line indicates scores of individuals not included in the figure. The within-cluster residual variance is assumed to be equal for all individuals (θ). B: Succinct representation of univariate random-intercepts model. The figure represents the model for a single generic individual (i) nested in cluster (j). The observed variable includes both person (i) and group (j) subscripts, whereas the cluster-level latent intercept and slope factors include the cluster subscript. Definition variables (X_{ij}) are represented by a diamond symbol. All model parameters are explicitly represented in the figure. C: Mplus representation of a random-slopes model. Within and between levels are separated by a dashed line. A dot on the within-regression indicates that the

and the latent factor covariances represent the cluster-level covariance between the intercept and slope factors:

$$\Psi_j = \begin{bmatrix} V(u_{1j}) & \\ CV(u_{2j}, u_{1j}) & V(u_{2j}) \end{bmatrix}.$$

Finally, the residual covariance matrix contains variances of residuals for all individuals in cluster j .

$$\Theta_j = \begin{bmatrix} \sigma_w^2 & & \\ & \sigma_w^2 & \\ & & \sigma_w^2 \end{bmatrix}.$$

Because individuals within a cluster are assumed to be interchangeable, residual variances are equal. For a cluster with N_j individuals, the residual covariance matrix is $\Theta = \mathbf{I}_{N_j} \otimes \sigma_w^2$, where \mathbf{I}_{N_j} is an identity matrix, and σ_w^2 is the common residual variance. As in the case of the random-intercepts model, the above matrices for the CFA formulation of the random-slopes model correspond to matrices of the corresponding MEM matrices: the factor-loading matrix (Λ_j) corresponds to \mathbf{X}_j and \mathbf{Z}_j matrices, the factor covariance matrix (Ψ) corresponds to the \mathbf{G}_j matrix, the residual-covariance matrix (Θ_j) corresponds to the \mathbf{R}_j matrix, and the latent factor mean vector is identical to the fixed-effects vector (β).

Model-Implied Covariance: Person-Specific Covariances

Given the previous formulation of a random-slopes model as a CFA model, the model-implied covariances and means are identical to the corresponding equations for the CFA formulation of the general MEM. Although the restricted CFA is used for modeling both random-intercepts and random-slopes model, the model-implied covariance matrix for the random-slopes model is very different from that for the random-intercepts model. This difference is due to the presence of X_{ij} as a definition variable and necessitates certain modifications to the underlying software before FIML estimation can be used to fit models with random slopes.

Recall that the model-implied covariance matrix for the random-intercepts model has a very simple structure with within-cluster covariance among any two Y_{ij} equal to the variance of the random intercept, and the variance for each individual is equal to the sum of the between-clusters and within-cluster variances. This is no longer true for the random-slopes model. The variance of Y_{mj} for individual m is

regression of Y on X is random at the between level. The within-level dependent variable is assumed to be random at the between level, which now becomes the random intercept. Between-level random intercept and slope are represented as latent variables.

$$V(Y_{mj}) = V(I) + 2X_{mj} * C(S, I) + (X_{mj})^2 V(S) + V(y^w),$$

and the covariance between scores of any two individuals Y_{mj} and Y_{nj} is

$$C(Y_{mj}, Y_{nj}) = V(I) + V(S)X_{mj}X_{nj} + C(S, I)X_{mj} + C(S, I)X_{nj}.$$

It is clear that for the random-slopes model, the covariance among individuals within a cluster is a function of person-specific values of X_{ij} . More simply, there is no single covariance matrix for the entire sample as in the case of conventional SEM. Instead, in order to fit a random-slopes model as an SEM model, covariances must be modeled as a function of predictors. As a result, conventional sample means- and covariance-based ML estimation is not suitable for modeling a random-slopes model. In principle, FIML estimation allows the covariance for each unit (in this case, the cluster) to be modeled as a function of an individual's values on predictors (in this case, X_{ij}). To do so, a special data-handling mechanism is needed that allows model parameters to be fixed to person-specific data values. Although FIML estimation has been available in popular SEM software for some time, the value of modeling covariances as a function of observed data values has not been widely recognized and hence incorporation of random slopes within SEM software has been slow in coming.

Random Slopes in ML-SEM: Empirical Example

We now illustrate a bivariate random-slopes model from a ML-SEM perspective. The data used in this example come from a bilingual (Spanish and English) research study involving 1,418 kindergartners clustered within 134 classrooms. Students were in one of two different types of language instructional programs: (a) immersion with an exclusive focus on instruction in English and (b) nonimmersion, including various levels and types of bilingual instruction. The dependent variables of interest include reading ability in English (RAE) and Spanish (RAS), and the independent variable is the level of blending words/Spanish (BWS). Blending words/Spanish is a component of phonological awareness and involves the ability to blend separate sounds to produce a word. It is hypothesized that the effect of BWS on RAE and RAS will vary across classrooms. In addition, the cluster-level variable, language instructional program, is hypothesized to predict the level of classroom RAE and RAS as well as the classroom variability in the effect of BWS on RAE and RAS, respectively. Within classrooms, the RAE and RAS residuals are expected to be correlated.

Table 5 and Figure 7 present the hypothesized model along with the parameter estimates. For each outcome (RAE and RAS), there are two classroom-level latent variables: (a) the random-intercept or the average level of the outcome at

Table 5
Bivariate Random-Slopes Model: Parameter Estimates

	Between-clusters model			
	English		Spanish	
	English intercept	English slope	Spanish intercept	Spanish slope
Latent variable covariance ^b				
Random factor				
English intercept	0.515 (.108)	.355	-.416	-.498
English slope	0.018 (0.011)	0.005 (0.002)	.210	.316
Spanish intercept	-.0242 (0.081)	0.012 (0.012)	0.656 (0.108)	.607
Spanish slope	-.0016 (0.009)	0.001 (0.001)	0.022 (0.011)	0.002 (0.001)
Fixed effects				
Predictor				
Intercept	-.0251 (0.085)	0.195 (0.013)	0.119 (0.097)	0.211 (0.010)
Immersion	1.332 (0.169)	0.049 (0.028)	-1.234 (0.167)	-0.076 (0.024)
Within-cluster model (residual covariance ^b)				
Residuals	English		Spanish	
English	1.593 (0.122)		.565	
Spanish	0.800 (0.070)		1.259 (0.092)	

Note. Correlations are above the diagonal. Model log likelihood = -4,350.045.

^a Within- and between-clusters covariances are conditional or residual covariance matrices.

^b Numbers in parentheses represent the standard error.

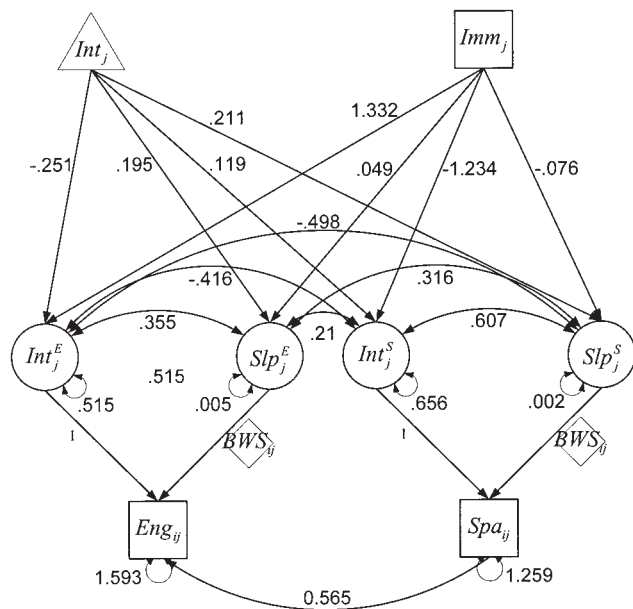


Figure 7. Parameter estimates for the bivariate random-slopes model for English (Eng) and Spanish (Spa) reading. Each outcome is represented as a separate random-slopes model as in Figure 6B. Intercept (Int) and slope (Slp) factors for each outcome (English & Spanish) are allowed to be correlated. Individual residuals for each outcome are also allowed to correlate. Parameter estimates are unstandardized except for interrelationships among variables at each level that are correlations. Imm = immersion; BWS = blending words/Spanish.

the grand mean of BWS and (b) the classroom-level variability in the effect of BWS on each outcome. For both outcomes, the factor loadings for the respective intercept factors are fixed to 1.0, and those for the slope factor are fixed to the person-specific values of BWS, represented within a diamond. The person-specific residuals for RAS and RAE are allowed to correlate. All four cluster-level latent variables are regressed on the dummy language program variable (coded: 1 = immersion; 0 = nonimmersion). As a result, the average Level 2 regression intercept for all four variables is defined for the nonimmersion group. The effect of the dummy language program variable on each latent factor represents the mean difference between the two program types.

Fixed Effects: Effect of Language Program on RAE and RAS Intercepts and Slopes

For the nonimmersion group, the expected levels of RAE intercept and slopes were -0.251 and 0.195 , respectively, and the expected levels of RAS intercept and slopes were 0.119 and 0.211 , respectively. The average intercept is determined entirely by the origin of the dependent variable and has little substantive meaning. Significant slopes for the nonimmersion group mean that

there was significant effect of BWS on both RAE and RAS for the nonimmersion group. The effect of the dummy language instruction variable on both RAE intercept (1.332) and RAS intercept (-1.234) was significant, suggesting that the two language programs were significantly different in the level of English and Spanish reading ability at the grand mean of BWS. As would be expected, the immersion group was significantly higher on the English intercept as compared with the nonimmersion group, whereas the nonimmersion was higher on the level of Spanish intercept. The effect of instructional language on RAE slope was nonsignificant (0.049 , *ns*), whereas the effect on the RAS slope was significant and negative (-0.076), suggesting that the average effect of BWS on Spanish reading ability was lower for the immersion group as compared with the nonimmersion group. In other words, the effect of BWS on Spanish reading ability is stronger for the group receiving some Spanish instruction as compared with the group receiving no Spanish instruction. On the other hand, BWS does not differentially influence English reading ability depending on the language program.

Between-Clusters Residual Covariance Among RAE and RAS Intercepts and Slopes

After controlling for the language of instruction, there was significant residual variability in RAE intercept (0.515) and RAS intercept (0.656) as well as RAS Slope factor (0.002). In addition, residual covariance between RAE and RAS intercepts was significant and negative (-0.242), suggesting that at the grand mean of BWS, classrooms higher in average English reading ability were lower in average Spanish reading ability. The remaining covariances among the cluster-level latent variables were not significantly different from zero. It was assumed that the covariance among the classroom-level random effects was the same across the two language program groups.

Within-Cluster Covariance Between RAE and RAS Residuals

At the individual level, there was significant residual variability in English and Spanish reading ability. In addition, English and Spanish language residuals were significantly and positively correlated ($.565$). The pattern of correlation between the reading abilities for individual residuals was the exact opposite of what was observed at the classroom level, in which the average reading ability intercepts were negatively correlated ($-.416$). This pattern of finding is appropriate in bilingual education and reflects the fact that the same outcome has a different meaning at the individual and classroom levels. At the individual level, the language abilities are expected to be positively correlated, reflecting the influence of individual-level factors such as intelligence and socioeconomic status; whereas at the class-

room level, the negative correlation between the average levels of reading ability reflects differences in the pattern of English and Spanish instruction across classrooms. Generally speaking, in the current sample, classrooms with higher levels of English instruction tended to have lower levels of Spanish instruction, and vice versa, explaining the negative correlation between the two abilities at the classroom levels.

Random Slopes Versus Cross-Level Latent Variable by Observed Variable Interaction

The empirical example presented above for illustrating the use of random slopes in SEM was rather trivial. However, the latent variable approach to modeling random effects offers an interesting alternative substantive interpretation more in line with the latent variable modeling perspective. From a MEM perspective, the effect of BWS on RAS and RAE is assumed to vary randomly across classrooms. This interpretation is implicit in the graphical representation used by Mplus (Figure 6C). From a latent variable perspective, we have classroom-level latent variables (random slope for RAS and RAE), perhaps indicative of an unobserved classroom-level construct, whose effect is moderated by the level of individual-level BWS. Both perspectives are consistent with the underlying mathematical model. However, the cross-level interaction between an observed and a latent variable is closer in spirit to the latent variable modeling perspective. This is clearly evident in the graphical representation of the model using the extended RAM notation (see Figure 7).

Discussion

The current article extended Mehta and West's (2000) approach of fitting LGC models for continuous time to the general case of ML-SEMs with random slopes. The article demonstrates the equivalence of the general multivariate MEM and ML-SEM with definition variables. MEMs are conceptualized as a special case of the restricted CFA model. The article also provides an extended RAM notation for graphical representation of ML-SEMs with random slopes. The extended graphical notation is a concise and mathematically complete representation of the model as well as its assumptions. The graphical notation should assist in conceptualizing complex ML-SEMs and also facilitate communication among researchers. The most important question from a practitioner's perspective is the following: How applicable are ML-SEM models for answering substantive questions of interest? In addition, in applied settings, methodological issues regarding the estimation and evaluation of ML-SEMs are critical. These issues are highlighted next.

Applicability of ML-SEM Models

MLM is generally thought of as multiple regression for clustered data. In the same way that SEM improves on multiple regression, ML-SEM brings the advantages of latent variable modeling to MLM for clustered data. Therefore, ML-SEM models are well suited to the kinds of questions for which SEM is best known.

Multilevel measurement models. The multilevel measurement model is perhaps the most important class of ML-SEM application. It is possible to estimate individual-level measurement models using observed or latent individual-level indicators. Similarly, it is possible to define cluster-level measurement models using the cluster-level random intercepts for the individual-level variables as well as cluster-level observed indicators. As in the context of conventional SEM, measurement modeling at each level of nesting opens up the possibility of investigations of measurement invariance. As illustrated in this article, we may be interested in the invariance of measurement model across levels of nesting. If across-level measurement invariance is justified, then an individual-level latent variable may be thought as having random variability at the cluster level. We may also be interested in across-group equivalence of measurement models. With clustered data, the grouping variable can be at the level of the individual (e.g., gender; see Curran, 2003) or at the level of the cluster (e.g., treatment administered at the level of the cluster). While a cluster-level grouping variable can be treated just as it would be in the conventional multiple group modeling, albeit for multilevel data, an individual-level grouping variable poses a bit of a challenge, as individuals belonging to different levels of the grouping variables are no longer independent within the same cluster. In such cases, we must now consider across-cluster variance and covariance across observed and latent variables for each grouping variable, in addition to the independent within-cluster variability for each group.

Multilevel mediation. The ML-SEM is well suited to evaluate the hypothesis of direct and mediated effects among latent variables at each level of nesting. ML-SEMs for latent mediation impose additional hypothesized restrictions on SEM models at each level of nesting. From a practitioner's perspective, this involves specifying additional directed relations among latent variables at each level (Kaplan & Elliot, 1997). At the level of the cluster, a model of mediation may involve latent variables that are defined using cluster-level indicators as well as those defined using individual-level indicators. For example, the latent variable of teacher quality may influence classroom-level teacher behavior, which in turn influences classroom variability in student-level outcome (e.g., student-level achievement at the classroom level).

Random slopes. As illustrated in this article, ML-SEM can also be used for estimating cluster-level variability in the effect of an individual-level variable on an individual

outcome. The dependent variable in this case may be an individual-level observed or latent variable. As in MLM and SEM, no distributional assumptions are made with respect to the strictly exogenous predictor other than the assumption of no measurement error.¹⁴ The cluster-level variable representing the random effect is in fact a latent variable and may be included as an independent or dependent variable in any cluster-level structural model. It is important to note that a model with a cluster-level random effect as a dependent variable of a cluster-level predictor (observed or latent) implies a hypothesis of an individual by cluster-level interaction between the cluster-level predictor and the individual-level variable whose effect is assumed to vary randomly at the higher level. In other words, an ordinary directional or mediational hypothesis involving a cluster-level predictor and a random effect for an individual-level predictor is actually a hypothesis of across-level moderation effect.

For naturally multivariate data, such as longitudinal or family research, it is possible to have random effects for individual-level variables that may also vary across clusters. For example, in individual growth curve models of students nested within classrooms, the effect of time on an outcome (observed or latent) is assumed to vary across individuals and also across classrooms. For univariate outcomes, within the MLM perspective, such models are conceptualized as three-level models (repeated measures within individuals and individuals within clusters). From a SEM perspective, such models may be treated as two-level models for multivariate data, provided that the data are time structured. If the time intervals between measurement occasions vary considerably, then it is no longer possible to use conventional missing data approach for fitting such models. Instead, it becomes necessary to model random effects of time using the definition variable approach (Mehta & West, 2000). This can be easily accomplished in software such as Mx.¹⁵

Evaluation of Model Fit

There is extensive methodological literature discussing evaluation of model fit in conventional SEM (e.g., Anderson & Gerbing, 1984; Browne & Cudeck, 1993), with simulation-based studies of alternative fit indices providing the bulk of the practical evidence (Bollen, 1986; Curran, West, & Finch, 1996; Gerbing & Anderson, 1993; Paxton, Curran, & Bollen, 2001). At a practical level, various fit indices attempt to answer two general types of questions: (a) Can we reject an ill-fitting and presumably incorrect model? and (b) Can we identify the "best" model from a set of competing and perhaps well-fitting models? The answer to these questions involves a reasoned argument regarding the power to detect a certain magnitude of disparity given the sample size, number of variables, and number of estimated parameters. Given the fact that ML-SEM is a variant of restricted CFA, the collective wisdom regarding evaluation of model fit in conventional SEM ought to be directly

generalizable to the case of ML-SEM. The nested data structure, however, adds an additional layer of complexity making the task a bit more challenging. Currently, very little is known regarding the properties of various fit indices in the context of ML-SEM. In some cases, even the definition of the fit index is in question. The following discussion attempts to identify relevant issues fully recognizing that the resolution of these issues will depend on future theoretical work and simulation studies.

Sample size(s) in ML-SEM. With nested data, there are at least two different sample sizes of interest: number of clusters (n) and total number of participants (N). Given the fact that ML-SEMs are really restricted CFA models fit to cluster-level covariances rather than the individual-level covariances, the appropriate sample size comparable to the conventional SEM sample size is the total number of clusters (n). The individual likelihood maximized in conventional SEM as well as in ML-SEM allows missing data. In ML-SEM, missing data includes both an unequal number of observations per individual and an unequal number of individuals per cluster (N_j), in which the maximum number of possible observation for a given cluster is $N_j * n_j$. The availability of different sample sizes raises questions regarding the appropriate sample size for computing various fit indices.

Likelihood ratio test (LRT) for nested-model fit. The fit of the base ML-CFA model is evaluated relative to the multivariate MEM by using an LRT comparing the difference in the fit functions ($-2LL$) given the difference in number of estimated parameters. In ML-SEM, the model is fit simultaneously to the between-clusters and within-cluster covariance. However, the likelihood is maximized with respect to the covariance defined at the level of the cluster. This has two implications: (a) The sample size in ML-SEM should refer to the total number of clusters (n) instead of the total number of individuals (N). (b) The fit of the model to the within-cluster covariance cannot be completely separated from the model fit to the between-clusters covariance.

The chi-square test of nested model fit is known to be sensitive to sample size (Bollen, 1986, 1990; Marsh et al., 1988). For the between-clusters CFA, this would be the total number of clusters. However, the effective number of observations may vary considerably across clusters. Hence, the between-level fit would also be a function of number of individuals per cluster (N_j). Similarly, the within-level model is not fit to individual data, but instead to the indi-

¹⁴ Models in which the random effect of a latent variable is of interest, alternative estimation methods involving integration are necessary (Skrondal & Rabe-Hesketh, 2004). Such methods are computationally intensive and are currently available in Mplus and GLLAMM.

¹⁵ At the time of writing this article, Mplus can fit such models using computationally intensive integration.

vidual-level residuals. As a result, it is likely that the chi-square test of fit for the within-cluster model would be influenced by the total number of individuals (N), the total number of clusters (n), and possibly the average number of individuals per cluster (\bar{N}_j) as well as the variability in the number of individuals per cluster.

Alternative fit indices. Recognizing that a large sample size can lead to rejection of trivially ill-fitting models, fit indices based on the principles of parsimony and information are becoming increasingly popular in SEM to identify appropriate models and to make decisions among competing models (root mean square error of approximation; Browne & Cudeck, 1993). Information-theoretic criteria (Akaike, 1973, 1987) are particularly useful in ML-SEM models with random slopes because the covariance-based fit indices are no longer meaningful. The definition and the role of these indices in ML-SEM is not clear, particularly with respect to the sample size to be used for computing these indices (i.e., N vs. n). This is most clearly seen with respect to the Bayesian information criterion (BIC; Schwarz, 1978) for which the role of the sample size and the number of estimated parameters (i.e., parsimony) is easy to understand. BIC is defined as

$$BIC = -2LL + r \ln N$$

where r is the number of free parameters and N is the sample size. The BIC is used for choosing the best model among a set of competing, non-nested models with different numbers of parameters with close but different values of the fit function. In such a situation, the LRT test is not suitable and the notion of information is invoked for comparing models. It is expected that a model with greater number of parameters will have a lower $-2LL$. The question then becomes the following: Are the additional degrees of freedom well spent? For the previous definition of BIC, a model with a smaller BIC is preferred on the grounds of parsimony. BIC penalizes a model with a greater number of estimated parameters. For two competing models, the model with more parameters will likely have a smaller value of $-2LL$ but will also have a larger value of $r \ln N$. Hence, a model with a slightly larger $-2LL$ may be selected if it has a sufficiently small number of parameters.

The value of N in the definition of BIC is a constant for any given sample and serves as an amplifier for the penalty. Everything else remaining the same, a larger sample size would result in a greater penalty ($r \ln N$) for the same number of parameters. For ML-SEM the question then becomes the following: What sample size (N or n) is appropriate for computing BIC? The choice of the sample size will have a nontrivial impact on the computed value of BIC, as the use of total sample size (N) would result in a far greater penalty as compared with the use of the number of clusters (n).¹⁶ Conceptually, the likelihood value for ML-SEM is computed by fitting the model at the level of the cluster and not

at the level of an individual; hence, the total number of Level 2 nesting units (n) would seem like an appropriate sample size. SAS Proc Mixed (SAS Institute, 1996) does in fact use n for computing BIC for multivariate-mixed effects model. In contrast, Mplus uses the total sample size (N) for computing BIC for an identical model. The two different formulas can result in different rank ordering of models and thus lead to different decisions regarding the preferred model. We expect that future theoretical and simulation work will clarify the issues involved in the choice of sample size for computing various fit indices as well as the utility of different fit indices for evaluating model fit in ML-SEM.

Estimation, Computation, and Software Implementation

Computational aspects of estimation, implementation of an algorithm in a specific software, and choice of model representation in scripts are related but distinct issues. In practice, however, these are often intertwined for the end user. There has been an almost exponential growth in the sophistication of software for fitting complex statistical models. This trend is likely to continue in the near future, and we expect that the specific issues discussed here are likely to be resolved before the current article reaches its intended audience.

The FIML estimation necessary for fitting ML-SEM models is currently available in other SEM software packages such as Mx (Neale et al., 2004) and Mplus 3.11 (Muthén & Muthén, 2004). At the current time, Mx has two major limitations: (a) Although Mx has the necessary feature set and the underlying optimizer is exceptionally fast, it is highly inefficient in the manner in which the multilevel likelihood is computed, limiting its use to models with a small number of individuals per cluster and a small number of variables. (b) Mx requires that the models are specified in terms of matrices and model-implied means and covariances. This places an undue burden on an end user whose sole interest is to estimate parameters as easily as possible. Mplus scripting language is easy to learn even for fitting complex ML-SEM models. However, at the time of writing, Mplus uses numerical integration for computing the likelihood for ML-SEM with random slopes at both levels. This

¹⁶ There are two parts to the BIC equation: $-2LL$, which is a constant, and $r \ln C$, in which $C = n$ or $C = N$. For a single model, the use of n or N does not change the rank order. For two competing models with different values of $-2LL$ (such that LL is a constant) and unequal number of free parameters, the choice of C can alter the rank order of the preferred model based on the value of BIC. This is because the different values of C differentially penalize or amplify lack of parsimony. For a sample with $n = 100$ and $N = 1,000$, amplification factors for n and N are 4.605 and 6.908, respectively. For the same absolute values of $-2LL$ and difference in $-2LL$, between two models, the BIC values for the two models may be different.

is inefficient and unnecessary for models with continuous outcomes as the likelihood can be computed analytically. As mentioned earlier, it is likely that these issues will be resolved in the near future.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csake (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317–332.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155–173.
- Arbuckle, J. A. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Hillsdale, NJ: Erlbaum.
- Bauer, D. J. (2003). Estimating multilevel linear models as structural models. *Journal of Educational and Behavioral Statistics*, 28, 135–167.
- Bentler, P. M., & Liang, J. (2003). Two-level mean and covariance structures: Maximum likelihood via and EM algorithm. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 53–70). Hillsdale, NJ: Erlbaum.
- Bollen, K. A. (1986). Sample size and Bentler and Bonett's non-normed fit index. *Psychometrika*, 51, 375–377.
- Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin*, 107, 256–259.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & K. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147–158.
- Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. (1988). *An introduction to HLM: Computer program and user's guide* (2nd ed.) Chicago, University of Chicago, Department of Education.
- Chou, C.-P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling*, 5, 247–266.
- Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, 38, 529–569.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory analysis. *Psychological Methods*, 1, 16–29.
- Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal design. *Psychological Methods*, 1, 236–248.
- du Toit, S., & du Toit, M. (2003). Multilevel structural equation modeling. In J. De Leeuw & I. G. G. Kreft (Eds.), *Handbook of quantitative multilevel analysis* (pp. 273–321). Boston: Kluwer.
- Eaves, L. J., Neale, M. C., & Maes, H. H. (1996). Multivariate multipoint linkage analysis of quantitative trait loci. *Behavior Genetics*, 26, 519–526.
- Foorman, B., Chen, D., Carlson, C., Moats, L., Francis, D., & Fletcher, J. (2003). The necessity of the alphabetic principle to phonemic awareness instruction. *Reading and Writing*, 16, 289–324.
- Fox, J.-P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, 66, 269–286.
- Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Newbury Park, CA: Sage.
- Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.
- Goldstein, H., & Browne, W. J. (2002). Multilevel factor analysis modelling using Markov Chain Monte Carlo (MCMC) estimation. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 225–243). Hillsdale, NJ: Erlbaum.
- Goldstein, H., & McDonald, R. P. (1987). A general model for the analysis of multilevel data. *Psychometrika*, 53, 455–467.
- Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407–434.
- Gustafsson, J.-E. (2002). Measurement from a hierarchical point of view. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 73–95). Hillsdale, NJ: Erlbaum.
- Harnqvist, K., Gustafsson, J. E., Muthén, B., & Nelson, G. (1994). Hierarchical models of ability at class and individual levels. *Intelligence*, 18, 165–187.
- Heck, R. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 89–127). Hillsdale, NJ: Erlbaum.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Hillsdale, NJ: Erlbaum.
- Kaplan, D., & Elliot, P. R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling*, 4, 1–24.
- Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36, 249–277.
- Longford, N. T., & Muthén, B. (1992). Factor analysis for clustered observations. *Psychometrika*, 57, 581–597.
- MacCallum, R. C., Kim, C., Malarkey, W. B., & Keicolt-Glaser,

- J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, 32, 215–253.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391–410.
- McArdle, J. J., & Anderson, E. T. (1990). Latent growth models for research on aging. In L. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 21–44). San Diego, CA: Academic Press.
- McArdle, J. J., & Boker, S. M. (1990). *RAMpath: Path diagram software*. Denver, CO: Data Transforms.
- McDonald, R. P. (1993). A general model for two level data with responses missing at random. *Psychometrika*, 58, 575–585.
- Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, P. W. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1–4. *Scientific Studies in Reading*, 9, 85–116.
- Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, 5, 23–43.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54, 557–585.
- Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28, 338–354.
- Muthén, B. (1994). Multilevel covariance structure analysis [Special issue] *Sociological Methods & Research*, 22, 376–398.
- Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological methodology* (pp. 453–480). Boston: Blackwell Publishers.
- Muthén, B., & Asparouhov, T. (2002). *Modeling of heteroscedastic measurement errors* (Version 1). Retrieved November 2003 from <http://www.statmodel.com/mplus/examples/webnote.html>
- Muthén, B., & Asparouhov, T. (2003). *Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in Mplus* (Version 1). Retrieved November 2003 from <http://www.statmodel.com/mplus/examples/webnote.html>
- Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (Ed.), *Sociological methodology* (pp. 216–316). Boston: Blackwell Publishers.
- Muthén, L. K., & Muthén, B. (2004). *Mplus user's guide* (3rd ed.). Los Angeles: Muthén & Muthén.
- Neale, M. C. (1998). Modeling interaction and nonlinear effects with Mx: A general approach. In G. Marcoulides & R. Schumacker (Eds.), *Interaction and non-linear effects in structural equation modeling* (pp. 43–61). Hillsdale, NJ: Erlbaum.
- Neale, M. C. (2000a). Flexible QTL mapping with Mx. In T. Spector, H. Snieder, & A. MacGregor (Eds.), *Advances in twin and sib pair analysis* (pp. 219–243). London: Greenwich Medical Media.
- Neale, M. C. (2000b). Individual fit, heterogeneity, and missing data in multigroup structural equation modeling. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples* (pp. 249–267). Hillsdale, NJ: Erlbaum.
- Neale, M. C. (2003). Quantitative genetics. In D. N. Cooper (Ed.), *Encyclopedia of the human genome* (pp. 679–683). London: Macmillan Publishers Ltd, Nature Publishing Group.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2004). *Mx: Statistical modeling* (5th ed.). Richmond: Virginia Commonwealth University, Department of Psychiatry.
- Paxton, P., Curran, P. J., & Bollen, K. A. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8, 287–312.
- Rabe-Hesketh, A., Pickles, A., & Skrondal, S. (2004). GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modeling Newsletter*, 13, 17–23.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, 69, 167–190.
- Rasbash, J., Browne, W. J., Goldstein, H., Yang, M., Plewis, I., Healy, M., et al. (2000). *A user's guide to MLwiN* (Version 2.1). London: University of London, Institute of Education.
- Raudenbush, S. W. (1995). Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. *British Journal of Mathematical and Statistical Psychology*, 48, 359–370.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Newbury Park, CA: Sage.
- Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model of studying school climate with estimation via the EM algorithm and application to U. S. high school data. *Journal of Educational Statistics*, 16, 296–330.
- Raudenbush, S. W., & Sampson, R. (1999). Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research*, 28, 123–153.
- Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8, 185–205.
- Rovine, M. J., & Molenaar, P. C. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research*, 35, 51–88.
- SAS Institute. (1996). *SAS/STAT software: Changes and enhancements through Release 6.11*. Cary, NC: Author.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Silberg, J. L., Rutter, M. L., Neale, M. C., & Eaves, L. J. (2001). Genetic moderation of environmental risk for depression and anxiety in adolescent girls. *British Journal of Psychiatry*, 179, 116–121.

- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hill.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Spiegelhalter, D., Thomas, A., & Best, N. G. (2000). *WinBugs Version 1.3 user manual*. Cambridge, England: MRC Biostatistics Unit.
- Thum, Y. M. (1997). Hierarchical linear models for multivariate outcomes. *Journal of Educational and Behavioral Statistics*, 22, 77–108.
- Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.
- Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, 116, 363–381.
- Yang, M., Goldstein, H., Browne, W., & Woodehouse, G. (2001). Multivariate multilevel analyses of examination results. *Journal of the Royal Statistical society*, 163, 49–62.

Received December 3, 2003

Revision received May 13, 2005

Accepted May 27, 2005 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://watson.apa.org/notify/> and you will be notified by e-mail when issues of interest to you become available!