

Journal of Educational and Behavioral Statistics

<http://jebs.aera.net>

Multilevel Growth Mixture Models for Classifying Groups

Gregory J. Palardy and Jeroen K. Vermunt

JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS 2010 35: 532

originally published online 14 September 2010

DOI: 10.3102/1076998610376895

The online version of this article can be found at:

<http://jeb.sagepub.com/content/35/5/532>

Published on behalf of



American Educational Research Association



<http://www.sagepublications.com>

Additional services and information for *Journal of Educational and Behavioral Statistics* can be found at:

Email Alerts: <http://jebs.aera.net/alerts>

Subscriptions: <http://jebs.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

>> [Version of Record](#) - Oct 14, 2010

[OnlineFirst Version of Record](#) - Sep 14, 2010

[What is This?](#)

Downloaded from <http://jebs.aera.net> at Tilburg University on September 15, 2013

Multilevel Growth Mixture Models for Classifying Groups

Gregory J. Palardy
University of California

Jeroen K. Vermunt
Tilburg University

This article introduces a multilevel growth mixture model (MGMM) for classifying both the individuals and the groups they are nested in. Nine variations of the general model are described that differ in terms of categorical and continuous latent variable specification within and between groups. An application in the context of school effectiveness research is presented. Schools are classified into three Type B effectiveness categories based on their mean student mathematics achievement growth trajectories, controlling for differences in students' backgrounds across schools. The classification outcome is regressed on a set of school practice variables to investigate the association between practices and cognitive development. Various issues related to model specification are discussed, including the use of covariates to identify substantively meaningful classes.

Keywords: *multilevel growth mixture model; latent class model; school effectiveness*

This article was motivated by a research problem involving the classification of groups (e.g., schools, neighborhoods, families, clinics, and firms) from mean individual growth trajectories in a multilevel and longitudinal design. The multilevel growth model with within-group mixtures (MGMM-W; Muthen, 2004) provides a strategy for classifying individuals who are nested in groups based on the properties of their growth trajectories, but does not classify groups. This article introduces an MGMM modeling framework that can be used to classify individuals based on their growth trajectories and to classify the groups they are nested in based on mean within-group trajectories. Our primary focus is on specifications of this model for classifying groups, which we refer to as the MGMM with between-group mixtures (MGMM-B).

The MGMM-B is applicable in the behavioral sciences, particularly education, where longitudinal and multilevel designs are increasingly common and group effects (e.g., school effects) on individual growth is the central interest

in the study. Some recent examples in the literature include school effects on student learning, family effects on child development, and neighborhood effects on change in individual psychological and behavioral outcomes. In each case, extending the multilevel growth model to include a model-based clustering of groups adds a new dimension to the research problem. Namely, it adds the capability of examining whether population heterogeneity in the between-group growth trajectories is due in part to the presence of unobserved subpopulations at the group level and, if so, what factors differentiate them.

We begin by reviewing the modeling developments that led to the MGMM-B, starting with a brief introduction to the literature on latent class and mixture models, growth mixture models, and the MGMM-W.¹ We then present an MGMM with classes at both the within-group and the between-group levels, which we conceptualize as an extension of the MGMM-W, but it can also be thought of as an extension of the multilevel latent class model (Vermunt, 2003). Next, we present the MGMM-B and illustrate it with an application in which schools are classified into Type B effectiveness categories (Raudenbush & Willms, 1995) and in which a school's class membership is predicted using measures of school practices. The complexity of this model necessitates several decisions during model specification, which we provide guidance on.

Background on Latent Class and Mixture Models

Although the terms “mixture model” and “latent class model” are sometimes used synonymously, they have distinct early literatures with latent class models focusing on categorical outcomes (Goodman, 1974; Lazarsfeld, 1950; Lazarsfeld & Henry, 1968) and finite mixture models focusing on continuous outcomes (Day, 1969; Wolfe, 1970). More recently, Aitkin (1999) and others proposed using mixture and latent class models for investigating whether the assumption of normally distributed random effects holds in random effects regression models. In the context of growth modeling, which is a specific type of random effects regression model, the recent literature has typically distinguished growth mixture models from latent class growth models, with the former carrying the assumption of normally distributed random effects within classes and the latter carrying the assumption that there is no random variation within classes (Bauer & Curran, 2004; Muthen, 2004). In that regard, the latent class growth model can be considered a special case of the growth mixture model where the random effects are set to zero within classes. It should, however, be noted that this distinction between mixture and latent class models is not used outside of the growth modeling context.

The distribution of random effects may depart from normal arbitrarily or because of the presence of subpopulations of individuals. Hence, latent class and mixture models may be used for two purposes. The first purpose is to test the normality assumption on certain random effects and provide a semiparametric or a nonparametric summary of the data when it is not met (Ferguson, 1983).

The second purpose is the substantive pursuit of “un-mixing” subpopulations of observations that differ in terms of the distributional properties of their random effects, such as mean or variance. That is, to classify observations into theoretically meaningful groups. It is the second purpose that applied researchers will likely find most attractive and is what motivates the application presented below. Unfortunately, it is difficult to verify whether non-normality is caused by mixtures of subpopulations outside of the artificial context of a simulation study (Bauer & Curran, 2003). Hence, when using these models with real data, the practitioner will generally be unable to ascertain whether the latent classes are detecting real subpopulations or non-normality that is an artifact of another origin.²

The growth mixture model and latent class growth model are recent extensions of the latent class and finite mixture model (Muthen, 2001; Muthen & Shedden, 1999; Nagin, 1999; Nagin & Tremblay, 2001; Verbeke & Lesaffre, 1996; Vermunt & Van Dijk, 2001). They represent a merger of the finite mixture model and the growth model and offer the advantages of each. These models overcome the assumption that random intercepts and slopes of individual growth trajectories be normally distributed and allow for the investigation of theoretically meaningful classes of individuals that differ in terms of their intercepts and slopes.

A Framework for MGMMs

A natural extension of the growth mixture model is the addition of a between-group level of analysis, which is useful for studying the association between group characteristics and the growth process when observations are nested in groups (e.g., schools or neighborhoods). The MGMM-W (Muthen, 2004) classifies individuals based on their growth trajectories and consequently may be considered an MGMM for within-group classification. The within-group latent class variable can be partitioned into within- and between-group components using a multilevel multinomial logistic regression model. The between-group component allows for the modeling of group factors on the probability of class membership for individuals. This model is useful for studying the associations between the within-group latent classes and group characteristics. However, another extension, the capability of basing latent classes on group intercept and slope estimates, is necessary for this modeling framework to realize its full potential for multilevel applications. This article introduces that extension as part of a multilevel latent variable modeling framework that can classify either within- or between-group observations, or both.³

Table 1 displays a matrix of nine potential models for describing growth trajectory heterogeneity. These models differ in terms of specification of continuous and discrete latent variables (i.e., latent classes) at both the individual and the group levels. The choice of specification can be important because the discrete and continuous latent variables compete for observed variation in the individual growth trajectories. In models that include latent classes, the decision of whether to retain the random effects can influence the number and composition of the classes that are

TABLE 1

Matrix of Nine Potential Multilevel Models for Describing Growth Trajectory Heterogeneity

Within-Group LVs	Between-Group LVs		
	Continuous	Discrete	Both
Continuous	A1	A2	A3
Discrete	B1	B2	B3
Both	C1	C2	C3

Note: LV = latent variables.

identified (Bauer & Curran, 2004; Lubke & Neale, 2006). There are three possible configurations at both the individual and the group levels including continuous latent variables only, discrete latent variables only, and both continuous and discrete latent variables. Because the nomenclatures can be a little convoluted, we connect all model names to letter–number identifiers to minimize confusion. Letters refer to the individual level and numbers refer to the group level. “A” and “1” annotate to models where all trajectory heterogeneity is captured by continuous random variables, namely, the random effects for the intercepts and slopes. “B” and “2” refer to models that capture heterogeneity using a set of discrete latent variables or latent classes. This specification implicitly assumes that the trajectories do not vary within class and hence there are no random effects for the slopes and intercepts. Note that model B2 is the multilevel extension of the latent class growth model (Nagin, 1999) and may be referred to as the multilevel latent class growth model (MLCGM). “C” and “3” are hybrids of the conventional growth model and latent class growth model (LCGM) that includes both latent classes and random effects for the intercepts and slopes. We refer to this model as the MGMM. At each level, the random effects may be specified within each latent class or only in the classes that have significant variation in the individual growth trajectories.

Model A1 is the conventional multilevel growth model, while models B1 and C1 are MLCGM-W and MGMM-W. The other six models, each of which identifies classes at the group level, are new.⁴ B2, B3, C2, and C3 draw latent classes from both levels. For these models, the group-level classes can originate from group intercepts and slopes and/or from the random effect associated with the individual-level latent class variable. The latter specification assumes that individual-level latent classes depend on the group-level classes and therefore may be considered a type of cross-level interaction effect, whereas the likelihood of being a member of a particular individual-level class depends partially on the between-group class the individual is a member of. The application presented in this article focuses on Models A2 and A3, which we refer to as the MLCGM-B and the MGMM-B, respectively. However, because C3 is the most complex

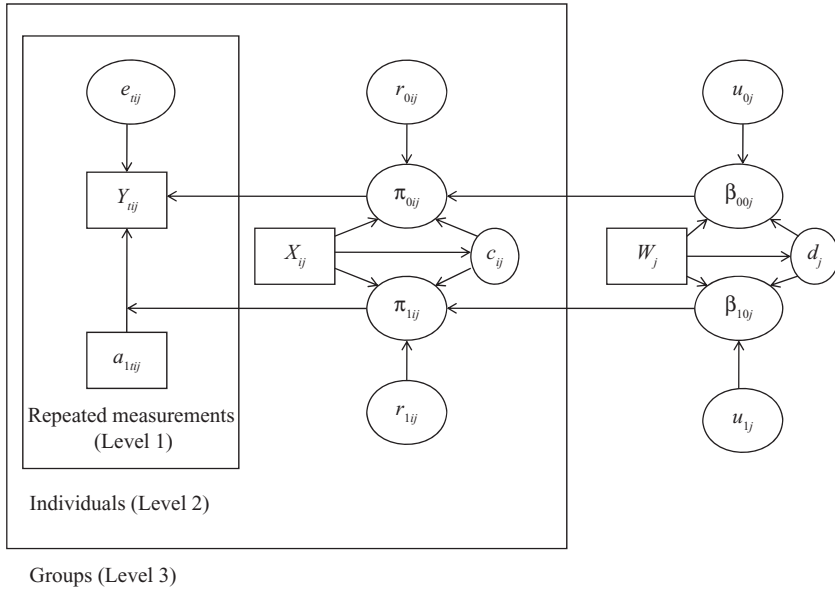


FIGURE 1. Path diagram of the multilevel growth mixture model.

model in Table 1, we present its path diagram and formulations. The other eight models can be considered special cases of C3.

Path Diagram

As a conceptual introduction to this modeling framework, we represent model C3 as a path diagram in Figure 1, which was inspired by the diagrams of Rabe-Hesketh, Skrondal, and Pickles (2004). Per structural equation modeling convention, rectangles represent observed variables and ovals or circles represent latent variables. Figure 1 shows the three-level nested structure of the data, with the repeated measurements (indexed t) nested within individual (indexed i), who are nested within groups (indexed j). Note that there are two kinds of latent variables represented: categorical, which are represented by circles, and continuous, which are represented by ovals. The categorical latent variables are the latent classes and are symbolized by c_{ij} at the individual level and by d_j at the group level, which are vectors of class indicators, where $1 \leq i \leq n_j$ and $1 \leq j \leq N$. The continuous latent variables are the random intercepts and slopes at the individual and group levels. At Level 1, we see the repeated measurements (Y_{ij}) are regressed on a measure of time (a_{1ij}) and there is a residual term associated with the repeated measurement outcome (e_{ij}), where $1 \leq t \leq T_{ij}$. The repeated measurements model generates growth trajectories for each individual, which are

summarized by the random intercept and slope coefficients (π_{0ij} and π_{1ij} , respectively). The individual-level intercepts and slopes may be regressed on a set of covariates (\mathbf{X}_{ij}) and have residuals associated with them (r_{0ij} and r_{1ij}). Moreover, latent classes (\mathbf{c}_{ij}) may be based on the individual trajectories. The mean individual intercepts and slopes for each group (β_{00j} and β_{10j}) are the outcomes of the group-level models. Classes (\mathbf{d}_j) can also be based on group trajectories, which may also be regressed on a set of group-level covariates (\mathbf{W}_j). The group-level models have residual terms associated with them (u_{0j} and u_{1j}).

Formulations

We now present model C3 in a series of five equations (1–3b). As with the path diagram, the formulations can be expressed in three parts including the repeated measurements model, the within-group model, and the between-group model. These equations extend the conventional three-level growth model equations (Raudenbush & Bryk, 2002) to include categorical latent variables that represent unobserved trajectory classes at the individual and group levels. We begin with Equation 1, which we refer to as the repeated measurements model. For simplicity, we present a linear trajectory form, but other functional forms can easily be used.

Repeated measurements model (Level 1):

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}a_{1tij} + e_{tij}. \quad e_{tij} \sim N(0, \sigma^2). \tag{1}$$

The notation in these equations uses subscripts to account for the multilevel data structure. Y_{tij} indicates the observed value of the outcome variable at time t for individual i nested in unit j (e.g., school), π_{0ij} is the expected value of Y for this individual when time equals zero, π_{1ij} is the expected change on the outcome per unit change in time for this individual, a_{1tij} is a variable measuring the passage of time within this individual, and e_{tij} denotes the residual or random error associated with this relationship. Whereas the presented linear growth model contains only one time variable, polynomial and piecewise linear growth models will contain multiple time variables. Hence, we use the subscript 1 to indicate this is the first of potentially multiple measures of time. The repeated measurements model generates intercepts (π_{0ij}) and slopes (π_{1ij}) for each individual, which are the outcomes for the within-group model, which we now turn our attention to. The within-group or individual model can be expressed in two parts including the model for the intercepts and slopes (2a) and the model for the classes (2b).

Within-group model for the intercepts and slopes (Level 2):

$$\begin{aligned} \pi_{0ij} &= \beta_{00j} + \sum_{k=2}^K \beta_{00k}^* c_{kij} + \sum_{q=1}^Q \beta_{0qj} X_{qij} + r_{0ij} \\ \pi_{1ij} &= \beta_{10j} + \sum_{k=2}^K \beta_{10k}^* c_{kij} + \sum_{q=1}^Q \beta_{1qj} X_{qij} + r_{1ij}, \quad \mathbf{r}_{ij} \sim N(\mathbf{0}, \mathbf{T}_r) \end{aligned} \tag{2a}$$

Within-group model for the classes:

$$\text{logit}[P(c_{kij} = 1)] = \lambda_{0k} + \sum_{q=1}^Q \lambda_{qk} X_{qij}. \tag{2b}$$

The model for the intercepts and slopes (Equation 2a) shows that variation in the individual growth trajectories is accounted for by three factors including covariates, latent classes, and random effects. Let X_{qij} be one of a set of Q covariates on which the intercepts and slopes are regressed to explain variation across individuals. The association between those covariates and the intercepts and slopes are represented by a set of Q coefficients, β_{01j} to β_{0qj} for the intercepts and β_{11j} to β_{1qj} for the slopes. To account for the addition of K latent classes, which are indexed by k with $1 \leq k \leq K$, let c_{kij} be one of K indicator variables taking on the value 1 if subject i is a member of latent class k and 0 otherwise. Note that only $K - 1$ of the K class indicators appears in Equation 2a because the class indicators are dummy coded with Class 1 being the reference category. Consequently, β_{00j} and β_{10j} represent the mean intercept and slope for Class 1, while β^*_{00k} and β^*_{10k} represent the difference between the mean intercept and slope for class k compared to Class 1. An alternative parameterization can be used that estimates the intercept and slope for each latent class, rather than the differences from the reference class. The equation for that parameterization will omit β_{00j} and β_{10j} from Equation 2a, which fixes them to zero, and index k to begin with Class 1 instead of 2.

Besides classifying individuals, the within-group model can be used to study the association between predictors and the likelihood of class membership. This part of the model, which is equivalent to a multinomial logistic regression model, is described in Equation 2b. Note that when $k = 1$ is the reference category, the left-hand side of the equation becomes $\log \frac{P(c_{ijk}=1)}{P(c_{ij1}=1)}$. Note that other types of logits could be used, for example, with a different reference category or with the geometric mean of the K categories in the denominator (effects coding). Moreover, this equation can be extended to a multilevel multinomial logistic regression model where the probability of class membership varies across between-group units.

The between-group model equations are highly similar to the within-group equations, as we show below in Equations 3a and 3b.

Between-group model for the intercepts and slopes (Level 3):

$$\begin{aligned} \beta_{0ij} &= \gamma_{000} + \sum_{l=2}^L \gamma^*_{000l} d_{lj} + \sum_{p=1}^P \gamma_{00p} W_{pj} + u_{0j} \\ \beta_{10j} &= \gamma_{100} + \sum_{l=2}^L \gamma^*_{100l} d_{lj} + \sum_{p=1}^P \gamma_{10p} W_{pj} + u_{1j}, \quad \mathbf{u}_j \sim N(\mathbf{0}, \mathbf{T}_u) \end{aligned} \tag{3b}$$

Between-group model for the classes:

$$\text{logit}[P(d_{ij} = 1)] = \delta_{0l} + \sum_{p=1}^P \delta_{pl} W_{pj}. \quad (3b)$$

As we saw with the within-group equations, heterogeneity in the trajectories is accounted for by three factors, which are now between-group covariates (W_{pj}), latent classes (d_{ij}), and random effects (u_{0j} and u_{1j}). The number of between-group latent classes is denoted by L and a particular between-group class by l , where $1 \leq l \leq L$. The indicator variable d_{ij} takes on the value 1 if between-group observation j belongs to between-group latent Class l and is 0 otherwise. The terms γ_{000} and γ_{100} appearing in Equation 3a are the mean intercept and slope trajectory for between-group Class 1, while γ^*_{000} and γ^*_{100l} indicate the differences in the mean intercepts and slopes for between-group Class l compared to between-group Class 1. Equation 3b describes the part of the model that predicts between-group class membership using covariates measuring group characteristics. Assuming $l = 1$ is the reference category, the left-hand side becomes $\log \frac{P(d_{ij}=1)}{P(d_{i1}=1)}$.

The residuals in the measurement, within-group, and between-group models are assumed to be (a) normally distributed, (b) independent across hierarchical levels, and (c) uncorrelated with the predictors included in the model. In standard regression analysis, the latter assumption is sometimes referred to as exogeneity of covariates. However, it should be noted that these three assumptions may be weaker in this multilevel mixture model compared with the standard multilevel model because they may be conditional on the class memberships (see extension 2 below).

Extensions

The model described in Equations 1–3b can be extended in various ways. Each of these extensions makes it possible to relax and test certain assumptions of the MGMM described above. The following are four potentially useful extensions:

1. Equation 1 assumes that the residuals in the repeated measurements model are uncorrelated and homoskedastic across time points, which may be untenable (Hedeker & Gibbons, 2006). To address this issue, a more flexible measurement error structure may be used. For example, the variance in the residuals in Equation 1 can be assumed to differ across time points (i.e., $e_{ij} \sim N(0, \sigma_i^2)$) or to have a full covariance structure (i.e., $e_{ij} \sim N(0, \Sigma)$).
2. In the above equations, it is assumed that residual variances do not differ across the classes. This restriction can be relaxed as it commonly is in applications of mixture models for continuous response variables (see McLachlan & Peel, 2000, pp. 81–83). In this multilevel model that would mean residual variances can be freely estimated across within-group classes, between-group classes, or both. That is, $e_{ij} \sim N(0, \sigma_k^2)$, $e_{ij} \sim N(0, \sigma_l^2)$, or $e_{ij} \sim N(0, \sigma_{kl}^2)$ for the residuals in the repeated

measurements model; $\mathbf{r}_{ij} \sim N(0, \mathbf{T}_{rk})$, $\mathbf{r}_{ij} \sim N(0, \mathbf{T}_{rl})$, or $\mathbf{r}_{ij} \sim N(0, \mathbf{T}_{rkl})$ for the residuals in the within-group model; and $\mathbf{u}_j \sim N(0, \mathbf{T}_{ul})$ for the residuals in the between-group level model.

3. Whereas in Equations 2a and 3a, only the mean intercepts and mean slopes are specified to differ across within- and between-group latent classes, the covariate effects in these equations can also be specified to differ across classes. In other words, the covariate effects can be specified to interact with the latent classes at either level by adding relevant product terms in these equations. That is, for within-group covariates by the within group classes, $\sum_{k=2}^K \beta_{0qk}^* X_{qij} c_{kij}$ for any $q = 1, \dots, Q$, where β_{0qk}^* is the coefficient on the interaction term; for the within-group covariates by the between-group classes, $\sum_{l=2}^L \beta_{0ql}^{**} X_{qij} d_{lj}$ for any $q = 1, \dots, Q$, where β_{0ql}^{**} is the coefficient on the interaction term; and for the between-group covariates by the between-group classes, $\sum_{l=2}^L \gamma_{0pl}^* W_{plj} d_{lj}$ for any $p = 1, \dots, P$, where γ_{0pl}^* is the coefficient on the interaction term. Note that similar terms but with an index of 1 instead of 0 will appear in the slope equations.
4. The logit parameters in the model for the within-group classes can be allowed to vary across groups by assuming that these are random effects (i.e., with a logit equation of the form $\text{logit}[P(c_{kij} = 1)] = \lambda_{0kj} + \sum_{q=1}^Q \lambda_{qkj} X_{qij}$ where $\lambda_{qkj} = \varphi_{qk} + v_{qkj}$ and $v_j \sim N(\mathbf{0}, \mathbf{T}_{vi})$). This multilevel extension was shown previously by Muthen (2004) for growth mixture models and by Vermunt (2003) for a general class of latent class and mixture models. An alternative strategy for specifying the within-class membership probabilities as group-specific is to include the between-group class indicators d_{jl} as predictors in the model for the within-group classes (see Vermunt, 2003). This implies including the term $\sum_{l=2}^L \lambda_{0kl}^* d_{lj}$ into Equation 2b and also possible interactions with the within-group predictors (i.e., $\int_{l=2}^L \lambda_{qkl}^* X_{qij} d_{lj}$ for any $q = 1, \dots, Q$).

Estimation and Software

Several options are available for the estimation of the unknown parameters of the MGMM described in Equations 1–3b. One option is to use Bayesian posterior estimation with either a Markov chain Monte Carlo (MCMC) algorithm implemented in a general purpose package like Winbugs (Spiegelhalter, Thomas, Best, & Lunn, 2003) or a tailor-made Gibbs sampler as, for example, Fox and Glas (2001) proposed for multilevel Item Response Theory (IRT) analysis. Another natural and feasible option is maximum likelihood estimation, which is the method we used in the application described below. More specifically, we used the maximum likelihood estimation procedure implemented in the Latent GOLD software

package by Vermunt and Magidson (2005, 2008). Appendix B provides more details on maximum likelihood estimation used by Latent GOLD, such as the form of the probability density function used to construct the log likelihood function.

Other software options for MGMM include latent variable modeling packages such as Mplus (Muthen & Muthen, 1998–2007) and GLLAMM (Rabe-Hesketh & Skrondal, 2008), which also use maximum likelihood estimation. One limitation of Mplus for MGMM is it uses a traditional latent growth curve approach with intercept and slope factors estimated using fixed factor loadings, which is suitable for repeated measurements that are balanced on time. As a result, Mplus currently cannot estimate the models shown in this article, which use a random time variable to estimate the intercept and slope factors and unbalanced repeated measurements. To be clear, the approach used in this article is suitable for either balanced or unbalanced repeated measurements. GLLAMM can estimate a subclass of the models introduced in this article; that is, the models with either only continuous random effects or only latent classes (our A1 and B2 model). Finally, rather than using a latent variable modeling package, some general statistical software packages such as R have specialized algorithms (e.g., *nlm* and *optim*) that may be used to optimize the log likelihood function to estimate the MGMM parameters.

Application: Classifying Schools Based on Student Cognitive Growth

We now turn our attention to an application of the MGMM for classifying schools into homogeneous groups based on the properties of their mean student math achievement growth trajectories. Although the MGMM framework introduced above includes nine models distinct in terms of specification of continuous and categorical latent variables, the context of the research problem of this application leads us to focus primarily on two of those that include a latent class variable only at the group level (i.e., Models A2 and A3 in Table 1). However, during the course of our analysis, we also estimate and discuss the multilevel growth model (A1) and an MGMM with latent class variables at both levels (B2). Similar to other complex models, specification decisions must be made that may influence the results, making the model specification strategy critical. Optimally, model building is not a purely statistical endeavor but rather is informed by the substantive literature and theory on the research problem (Nagin, 2005). For this reason, we begin by introducing a theoretical framework for estimating school effects, which guides the example. We then describe several model specification considerations that can be important for ensuring sensible results. But first we describe our data.

Data

This application uses a sample of 3,401 children nested in 254 schools from the Early Childhood Longitudinal Study of the Kindergarten Class of 1998–1999 (ECLS-K; National Center for Education Statistics [NCES], 2002).⁵ The

outcome is four repeated measurements of mathematics achievement scores collected near the beginning and end of kindergarten and first grade, which are used to estimate mathematics achievement growth during kindergarten and first grade. The test consists of a pool of items developed by National Center for Education Statistics (NCES) and set to a common scale using IRT. Test versions of varying difficulty were constructed and vertically equated. Each student at each data collection point was administered a test of appropriate difficulty for their level of cognitive development in mathematics. This longitudinal test development protocol was for the purpose of producing achievement data suitable for growth modeling. Besides repeated measurements of mathematics achievement, we also use student background variables measuring ethnicity and socioeconomic status (SES) as well as mean student SES at the school level. These variables are conceptualized as Type B control variables, for the purpose of which we describe below. Finally, we also use two variables measuring school practices: a dummy coded variable indicating whether school-based management (SBM) was being used and a factor score measuring a construct we refer to as Teacher Professionalism. We elaborate on these two measures of school practices below. In Appendix A, Table A1 shows the descriptive statistics for the variables used in this analysis and Table A2 shows the details on the Teacher Professionalism factor score.

Model Building

In this section, we introduce and provide guidance on several issues related to model building, including the theoretical framework informing model specification, the form of the growth curve, determining the number of classes, and potential misspecifications of the model used to classify the schools, which we refer to as the classification model. These issues are important because model specification decisions affect the number and composition of the classes that are identified.

Theoretical Framework Informing Model Specification: Type B School Effects⁶

This study seeks to classify schools into what we call Type B effectiveness categories based on mean student learning trajectories. Raudenbush and Willms (1995) proposed a conceptual framework for school effects that makes the distinction between Type A effects, the total impact that schools have on a student outcome such as learning, and Type B effects, the impact that school practices have on the student outcomes. Public school personnel typically have little control over the background composition of the students who enroll in their schools or certain other school features such as the level of resources that are available (e.g., per pupil expenditures) or the structural features of the school (e.g., enrollment or sector), all of which have been found to affect student learning. However, teachers have a fair amount of control over instructional practices and administrators have control over management practices, which also affect

student learning. Type A effects include both factors within and outside the control of school personnel and is typically what parents are most interested in when judging the effectiveness of schools. Type B effects describe the effectiveness of the school controlling for factors beyond the influence of school personnel and is what school evaluators are typically most interested in when judging the effectiveness of the school. For a learning outcome, the Type B effect for each school is the degree to which the school's mean achievement growth trajectory deviates from what is expected, controlling for student background and school factors that are beyond the influence of school personnel.

This paradigm of school effects is useful for informing model specification in an effort to produce more substantively meaningful results. When statistical adjustments are made for the effects of other factors, the Type B effects model provides a more appropriate foundation for comparing the effectiveness of schools based on the performance of school site personnel. In this application, we use a Type B school effects model to estimate the optimal number of between-group trajectory classes. The model controls for aspects of students' background that affect learning as well as nonpractice school factors. In the within-school model, we control for the student SES as well as ethnicity, two demographic variables that have shown robust associations with learning in early elementary school. In the between-group model, we control the social composition of the student body, which is measured by the mean SES of the school.⁷

The estimate of Type B school effects by Raudenbush and Willms is based on between-group slope random effects or residuals. The current study uses the MGMM to classify schools into Type B effectiveness groups based on mean student achievement growth trajectories. There are a few noteworthy differences in these two approaches for estimating Type B school effects. First, the Raudenbush and Willms approach does not use the school intercepts (initial mean student achievement) directly when assessing effectiveness, which can be a confounding factor in the Type B effectiveness paradigm because the achievement level of entering students is beyond the control of school site personnel and it may be correlated with subsequent growth in achievement. In contrast, the classification of schools into homogeneous groups by the MGMM is based on both the school intercept and the slope, which addresses this concern. Moreover, the mean intercept and slope covariances may differ across school classes. The MGMM approach also addresses the important practical issue associated with the continuous approach of determining which schools are performing well and which are not. That determination requires designating cut-points in the distribution of school slope residuals that discriminate effective and ineffective schools. Such cut-points are typically set based on an arbitrary criterion. For example, schools with slope residuals one standard deviation above average may be deemed effective. The use of such an arbitrary criterion is unnecessary with the MGMM approach.

Form of the Growth Curve: Piecewise Linear Growth Model

One of the strength of growth modeling is that with the appropriate data collection design, the time variables can be coded to provide more sensitive estimates of the treatment effect of interests. In the current example where the growth trajectories include multiple school years and achievement data were collected at the beginning and end of each school year, a piecewise linear growth model can be used, which partitions the summer period away from the school year when the schooling “treatment” is being administered. This specification of the time covariates provides a more sensitive estimate of the schooling treatment effect because the learning that takes place over the summer period when school is not in session generally cannot be attributed to school effects and also because learning rates for socially disadvantaged children (i.e., children from low SES families) tend to be lower during summer compared with their more affluent counterparts, although similar during the school year (Alexander & Entwisle, 1996; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). Hence, failing to separate the summer period from when school is in session may introduce a substantial source of measurement error to school year achievement growth estimates, which will tend to result in the underestimation of the effectiveness of schools serving disproportionately high percentages of disadvantaged children and overestimation of the effectiveness of schools serving affluent families.

The piecewise linear trajectories can take on various configurations. For example, the trajectories may be partitioned into three pieces including kindergarten, summer, and first grade, which produce separate estimates of the learning rates during each period. Figure 2 shows the mean student math achievement growth trajectories for each period for our sample of the ECLS data, controlling for student SES, ethnicity, and mean SES. The graph illustrates that the mean math learning rate during the kindergarten and first grade periods are similar, while the rate for the summer piece is substantially lower. Based on this observation, a piecewise linear growth model was used, which estimates the rate of achievement growth for the school year and summer separately.

Also in an effort to minimize measurement error, the model was specified to take into account two aspects of the ECLS data set testing protocol. The fall and spring achievement test scores were not collected at the very beginning and end of each school year. Moreover, the time between collections varied considerably across students (i.e., the data were not balanced on time).⁸ Two continuous “time” variables, one for the school year and one for the summer, were constructed to account for these design issues. The time variable for school year learning measures the passage of time in months since the start of kindergarten. The time variable for summer is coded zero from the start of kindergarten through the start of summer and in months during summer and is held constant from the end of summer until the end of first grade. Table 2 provides data for the

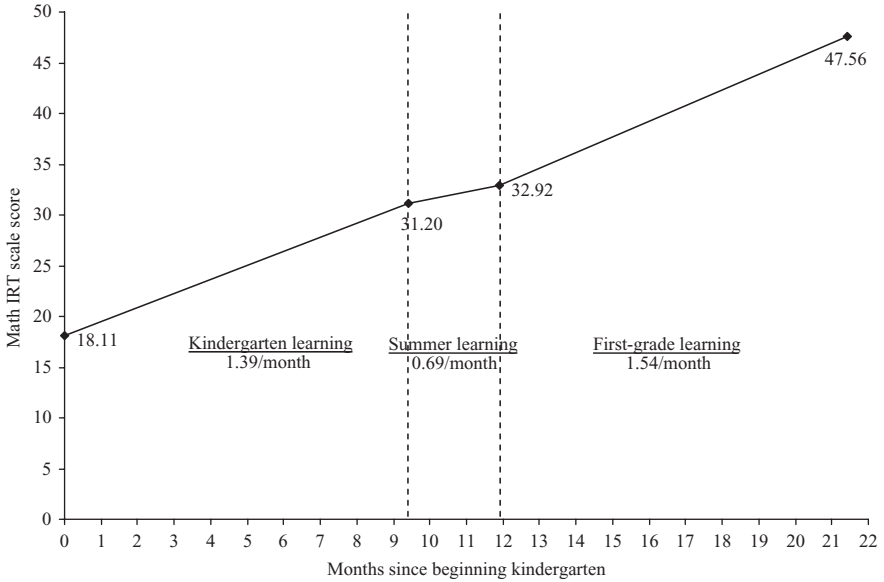


FIGURE 2. Mean piecewise linear growth trajectory for math achievement for Early Childhood Longitudinal Study (ECLS) sample. Note: Estimated values are based on conditional piecewise growth model, controlling for student background and social composition.

TABLE 2
Time Variable Coding Example: Data for Two Students on the School Year and Summer Time Variables

Student #	Repeated Measurement	a_{1tij} (School Year)	a_{2tij} (Summer)
1	1	1.45	0.00
1	2	8.55	0.00
1	3	12.85	2.55
1	4	19.03	2.55
2	1	2.01	0.00
2	2	9.07	0.00
2	3	14.04	3.02
2	4	20.48	3.02

school year and summer time variables, a_{1ti} and a_{2ti} , for two students in our sample. Note that both time variables are unbalanced, making these data unsuitable for the traditional latent growth curve (LGC) approach because LGCs use intercept and slope factors with fixed loadings, which require that the repeated

measurements be balanced on time (Bollen & Curran, 2006). This coding scheme results in a school year slope that can be interpreted as the expected monthly rate of math achievement growth while school is in session (adjusted for the summer period) and the summer slope that is the expected difference in the summer development rate compared with the school year rate. The beauty of this repeated measurements model is that it separates school year learning on which schools will be classified into effectiveness categories, from summer learning, which will not be used in the classification model. Whether estimating Type A or Type B school effect, this is an important specification issue.

Indices for Determining the Number of Classes

Several statistical indices are available for determining the number of classes in mixture and latent class models and authors have generally advised using more than one (Bauer & Curran, 2004; McLachlan & Peel, 2000; Nagin, 2005). In this example, we use four different information criteria including, perhaps, the two most commonly used, the Bayesian information criterion (BIC; Schwarz, 1978) and the Akaike information criterion (AIC; Akaike, 1973).⁹ These are comparative indices that are appropriate for determining which set of models fits the observed data best. Both indices are based on the model log likelihood and penalize for increased model complexity. BIC also penalizes for increased sample size. The formula for BIC is

$$\text{BIC} = -2 \log L + \log(N) \cdot \text{Npar}, \quad (6)$$

where $\log L$ is the model log likelihood, N is the number of observations, and Npar is the number of parameters for the fitted model. The formula for AIC is

$$\text{AIC} = -2 \log L + 2 \cdot \text{Npar}. \quad (7)$$

AIC penalizes for model complexity by adding twice the number of parameters. Note that the complexity adjustment for BIC becomes progressively larger as sample size increases and will be larger than the adjustment for AIC for sample sizes greater than 7. One potential strength of BIC for assessing multilevel mixture models is that the number of observations used in the penalty can be measured in various ways with perhaps the two most obvious choices being the number of individuals and the number of groups. This choice can affect the sample size penalty substantially, which in turn can affect the number of classes identified. When the model only classifies groups, the parameters differentiating comparison models may only be estimated from the groups rather than the within-group units, in which case the number of groups would seem the appropriate N for the BIC sample size adjustment. Supporting this rationale, a recent simulation study found that BIC with N equal to the number of group (which we refer to as BIC-B as opposed to BIC-W for which N is the number of individual observations) outperforms other indices for detecting the correct number of

between-group classes for a multilevel mixture model (Lukociene & Vermunt, 2010).¹⁰ To determine the optimal number of classes in the application below, we use BIC-W, BIC-B, and AIC, as well as the modified AIC (AIC3; Bozdogan, 1993), which research suggests performs well on a variety of latent class models (Dias, 2006, 2007). AIC3 differs from AIC only in that the penalty constant is three times the number of parameters rather than two times.

Potential Misspecifications of the Classification Model

In this section, we examine potential misspecifications of the classification model, which we define as the model that is used to determine the number of latent classes. We focus on three aspects of model specification that have potential for affecting the number and composition of the between-group classes we seek to identify (a) covariates, (b) between-group random effects, and (c) within-group latent classes.

The way classes differ from one another may change considerably when covariates are entered into the model (Muthen, 2004; Tofighi & Enders, 2007). This implies that the assignment of schools to classes based on their posterior membership probabilities may also be altered. Recall that covariates adjust the intercepts and slopes. Recall also that the degree of the adjustment for a particular unit is a function of the unit's values on the covariates and the outcome. Hence, the use of covariates in the classification model alters the distribution of the random effects and may change the composition and number of classes. For these reasons, it is important to carefully consider the use of covariates in the classification model. As suggested above, covariates should be selected based on the theoretical framework guiding the study. In this application, we include covariates that adjust for student and school characteristics that are arguably beyond the control of school site personnel because the objective is to classify schools into Type B effectiveness categories. Had the objective been to classify schools into Type A effectiveness categories, the choice of covariates in the classification model would have changed accordingly.

Once the covariates for the classification model have been determined, it must be decided whether their effects are fixed across classes or interact with the classification variable. If their effects vary across classes, misspecifying them as fixed could bias their adjustments and in turn affect the definition of the between-group classes. Consequently, we recommend testing whether the effects of covariates in the classification model vary across classes, which we refer to as a covariate invariance test. In summary, there is no absolute correct specification of covariates in the classification model. The decisions of whether to use covariates in the classification model, and if so which covariates, is best informed by the theoretical framework and objectives of a given study.

A second specification issue involves the conditions under which to leave the intercept and slope random effects unconstrained at the group level. Recent

simulation studies have shown that constraining these parameters to zero can result in overextraction of classes (Bauer & Curran, 2004; Lubke & Neale, 2006). Overextraction can also be caused by non-normality in the intercept and slope distributions that does not originate from mixtures of subpopulations (Bauer & Curran, 2004). A simple strategy for protecting against overextraction is to retain the random effects. However, in the context of the MGMM, the intercepts and slopes tend to vary far more within groups than between groups and the variance in the group slopes is usually several times smaller than the variance in the group intercepts (more than 300 times smaller in the example below). Hence, the between-group slope variance tends to be very small, which in our experience often results in a drastic slowing of parameter estimation and sometimes non-convergence. In such cases, it may be better to constrain the variance in the group slopes to zero. If this is done, a sensitivity analysis is recommended to examining whether the constraint impacts the number and composition of classes identified.

A third specification issue is whether to include a within-group latent class variable. When the objective is to classify groups and there is no substantive interest in classifying individuals, omitting the within-group latent class variable is reasonable and will result in a more parsimonious model. Yet, doing so may violate the assumption that the intercept and slope random effects are normally distributed and individuals within the groups are from a single population. Unfortunately, little is known about how such violations affect between-group classes. One thing that is clear, however, is that these models are very complex and prone to estimation challenges. Although some researchers may wish to pursue various extensions to the models, adding complexity that is not consistent with addressing the research questions is not advisable. Consequently, we recommend adding within-group latent classes only when the research problem calls for them.

Results

We present the models in order of progressive complexity. Although our theoretical framework calls for using a Type B school effects model to determine the number of classes, we also present the unconditional classification model results for comparison (see Table 3). We begin with the multilevel random coefficient growth model (MRCGM; A1), which is the conventional multilevel growth model and has only one class. We then estimate MGMM (A2), in which the between-group random effects are set to zero, starting with a single class and adding class one at a time until the information criteria indices increase, indicating a decline in model fit to the data. Next, we examine the MGMM (A3) with the between-group random effect for the intercept free and constrained to zero for the slope. We again start with the single-class model and add classes one at a time.

TABLE 3

Determining the Optimal Number of Classes for the Unconditional and Type B School Models

#Classes	Unconditional				Type B			
	BIC-W	BIC-B	AIC	AIC3	BIC-W	BIC-B	AIC	AIC3
Multilevel random coefficient growth model (A1)								
1	84,571	84,548	84,516	84,525	<u>84,195</u>	84,128	84,036	84,062
MGMM with between-group fixed slopes and intercepts (A2)								
1	85,202	85,184	85,159	85,166	84,309	84,249	84,168	84,191
2	84,749	84,723	84,688	84,698	84,245	84,178	84,086	84,112
3	84,656	84,622	84,576	84,589	<u>84,202</u>	<u>84,126</u>	<u>84,024</u>	<u>84,053</u>
4	84,614	84,573	84,516	84,532	84,225	84,142	84,029	84,061
5	84,605	84,556	84,489	84,508	—	—	—	—
6	84,617	84,560	84,482	84,504	—	—	—	—
MGMM with between-group random intercepts and fixed slopes (A3)								
1	84,663	84,642	84,614	84,622	84,258	84,195	84,111	84,135
2	84,583	84,555	84,516	84,527	84,206	84,135	84,040	84,067
3	84,577	84,541	84,491	84,505	84,220	84,142	84,036	84,066
4	84,594	84,550	84,490	84,507	—	—	—	—

Note. Bold numbers indicate best fitting model of a particular type, whereas underlined numbers indicate best fitting model overall. AIC = Akaike information criterion; AIC3 = modified Akaike information criterion; BIC = Bayesian information criterion; MGMM = multilevel growth mixture model.

Determining the Number of Classes

The results in Table 3 show that the optimal number of between-group classes differed slightly depending on the information criteria, which should be expected, given the differences in the formulations of the criteria. Among Unconditional models, the five-class and six-class models with the group intercepts and slopes variances constrained to zero (A2) each fit best for two of the four indices, while the three-class MGMM with a random intercept (A3) fits best for three of the four indices.¹¹ This finding corroborates previous research indicating that omitting random effects can lead to overextraction (Bauer & Curran, 2004; Lubke & Neale, 2006). However, adding the Type B covariates seems to alleviate that condition. The Type B classification models fit the data systematically better than the Unconditional models, with the three-class model with fixed between-group intercepts and slopes (A2) fitting best overall based on three of the four indices. This is likely a result of the Type B model covariates having explained a substantial proportion of the variance in the intercepts among schools, consequentially the between-group random effect for the intercept was no longer necessary for preventing overextraction. Finally, we estimated the Type B model with a random intercept (A3). The two-class and three-class models each fit best on two of the four indices.

TABLE 4
Variance Components for Unconditional and Type B Models

	Unconditional	Type B	
	Variance	Variance	% Variance Explained
Within intercept	37.4509	35.0325	6.46
Within slope	0.0377	0.0373	1.06
Between intercept	14.8439	3.3916	77.15
Between slope	0.0114	0.0093	18.42
ICC intercept	0.1454	0.0883	—
ICC slope	0.2013	0.1996	—

Note: ICC = intraclass correlation coefficient.

TABLE 5
Performance of Type B Classification Model: Cross Tabulation of Modal Class Assignments and Probabilistic Class Assignments

Probabilistic	Modal			Total
	1	2	3	
1	127.48	7.13	9.71	144.32
2	9.03	39.02	2.34	50.39
3	11.49	2.85	44.95	59.29
Total	148.00	49.00	57.00	254.00

However, adding the intercept random effect to the Type B model resulted in a systematic increase in the information criteria. Therefore, that specification was rejected in favor of the more parsimonious model with fixed intercepts and slopes at the group level.

To illustrate the effects of adding the Type B control covariates to the classification model, Table 4 presents the variance components for the MRCGM (A1) for the Unconditional and Type B classification models. We see that the covariates in the Type B model account for more than 77% of the between-group variation in the intercepts. These factors were controlled to account for variance in the school growth trajectories due to differences in Type B control variables. One impact this has on the model is that the intercept random effect no longer seems necessary for preventing overextraction.

Table 5 provides a sense of the performance of the Type B classification model. Schools were classified into their most likely (modal) class based on their posterior class membership probabilities for which the frequencies are shown in the marginal row at the bottom. These membership probabilities are

aggregated for the schools assigned to Classes 1, 2, and 3, respectively, for which the frequencies are shown in the marginal column on the right. The cell entries tabulate the frequency of consistency in the two methods of classification. If the model performs perfectly based on consistency, the off diagonal frequencies will all be zero and the marginal frequencies for the modal and probabilistic classifications schemes will be equal. The estimate of the proportion of misclassifications is obtained by dividing the sum of the off-diagonal elements by the sample size (Goodman, 1974). The estimate for our classification model (0.17) suggests that, although imperfect, 83% of the schools in the sample are correctly classified.

In summary, our analysis of these various classification models provides considerable support for the three-class model. Moreover, once the Type B control covariates are in the model, the group-level random intercepts are no longer necessary. Therefore, we retain the three-class Type B control model with fixed intercepts and slopes (A2), which we will henceforth refer to as the classification model.

Examining Potential Sources of Model Misspecification

Before investigating the relationship between class membership and school practice predictor variables, we examine the sensitivity of the model to three alternative specifications. The objective is to examine how the model fit, the number of classes identified, and the composition of the classes may be altered by these changes in the specification of the classification model. The first alternative specification is extension 3 above, which includes interactions between covariates and the classes. The second includes within-group classes. The third is the unconditional model, which contains no covariates.

Once the covariates are in the classification model, it is a sound practice to test whether they interact with the between-group classes. At issue here is the biasing of the between-group class-specific mean slopes and intercepts as well as the possible misclassification of groups. To address this, we conduct a test of covariate invariance across the classes. Table 6 compares the intercept and slope estimate for each class and the proportion of schools in each class for the classification model and the model with covariate–class interactions. Note that the information criteria for the interaction model were markedly higher (e.g., BIC-B = 84,198) compared with the classification model, suggesting an insufficient improvement in model fit to justify the addition of the interaction effect.¹² Moreover, the classes for either model were highly similar in terms of intercepts, slopes, and percentage of sample in each class. Only a very small percentage of the schools changed classes when the interaction was added. Given these findings, we retain the more parsimonious Type B effects model over the interaction model.

TABLE 6
Comparison Class Intercepts, Slopes, and Percentage of Sample for Alternative Classification Models With the Type B School Effects Model

Class	Type B Effects ^a			Interaction			Within-Group Class			Unconditional		
	I	S	%	I	S	%	I	S	%	I	S	%
1	16.09	1.57	56.82	15.96	1.57	56.94	16.17	1.57	59.22	19.06	1.49	47.66
2	15.43	1.35	19.84	14.80	1.37	16.81	15.33	1.35	16.91	26.12	1.35	5.77
3	20.03	1.43	23.34	20.09	1.42	26.24	19.90	1.42	23.87	13.13	1.44	46.57

Note. I = intercept; S = slope; % = percentage of sample.

^a Used to classify the schools.

We also examined the impact of adding two within-group latent classes on the between-group classes. This model fits the data better than the classification model (BIC-B = 84,001), indicating non-normality in the within-group random effects. However, the between-group classes hardly change terms of their intercepts and slopes nor does the percentage of observations in each class when the within-group classes are added (see Table 6). Moreover, as we saw for the interaction model, only a small percentage of the schools change classes. These findings suggest that in the current application, the between-group classification model is robust to potential violations of the normality assumption at the within-group level. Alternative specifications such as adding interactions between the classes and covariates or adding within-group classes have little impact on the composition, slopes, and intercepts of the between-group classes.

The classification model includes Type B control covariates, per the conceptual framework. To examine how the covariates effect the composition of the classes, we estimated an Unconditional model (see Table 6) that differs from the classification model only in that no covariates are included. The difference in the classes compared to the previous three models is rather striking. The composition of the classes is clearly different as the percentage of schools in each class has changed substantially. Now there is one small class of schools (5.77% of the sample) in which students enter kindergarten with math achievement levels roughly four grade levels above average but learn at a far slower rate than average. The other two classes are split nearly equally in size with one having students who enter kindergarten with high-average math achievement and learn fast and the other having students who enter with very low math achievement and learn at a moderate rate. Adding the Type B covariates to this model alters the school trajectories enough so that a fair percentage of the schools change classes. Indeed, the best Unconditional model has five or six classes, so it is more accurate to say that the school trajectories were redistributed with the addition of covariates resulting in fewer classes. The addition of the Type B controls was

TABLE 7
Results of the Multinomial Model for the Classes

Class	Class 1 ^a	Class 2	Class 3
Description	Effective/ Moderate	Very Ineffective/ Low	Ineffective/ High
Trajectory properties			
Intercept	16.53	15.87	20.69
Slope	1.57	1.35	1.40
Percentage of sample	58.57	21.97	19.46
Predictors of class membership ^b			
Constant (intercept)	—	0.30**	0.28**
Professionalism	—	0.35**	0.51*
School-based management	—	0.97	1.67

Note. ^aReference category; ^bScaled in odds metric; *Significant at .05; **Significant at .01.

expected to adjust the school trajectories. Indeed, that is precisely why they are included in the classification model, as outlined in our conceptual framework.

Description of the Type B Effectiveness Classes

We have determined that the three-class model fits the data best. Now we turn our attention to describing those Type B effectiveness classes. The upper panel of Table 7 shows that the class intercepts range from 15.87 to 20.69, whereas the slopes range from 1.35 to 1.57. Although the greater range in the intercepts compared with the slopes may seem to imply the intercepts are more influential for distinguishing the classes, that is not the case. The range for the intercepts is 4.82 points, whereas the range for the slopes is 0.22, which results in a cumulative slope effect of 4.18 ($.22 \times 19$) over the approximately 19 months duration of the 2 school years being modeled.

The intercept and slope results suggest that Class 1 schools are effective because students attending those schools tend to enter kindergarten with slightly below average math achievement (16.53 compared with 17.25 for average) and learn at a greater rate than the population average (1.57 vs. 1.49). We describe Class 1 schools as “Effective/Moderate” because they are effective in terms of mean learning rates and, on average, students enter with moderate achievements levels. Class 2 and 3 schools are ineffective because students attending schools in those classes tend to learn at substantially lower rates than average (1.35 and 1.40, respectively). Those two classes differ primarily in their intercepts. Student attending Class 2 schools tend to enter with below average math achievement (15.87), while students attending Class 3 schools enter with far

higher than average math achievement (20.69). We therefore label Classes 2 and 3 as Very Ineffective/Low and Ineffective/High, respectively.

Predicting Class Membership

The classification model categorizes schools based on their mean learning trajectories into three homogeneous classes. Now the objective is to differentiate these effectiveness classes of schools using practices that school personnel have control over. In this application, we examine two group-level variables that are hypothesized to be associated with the odds of class membership. One is a standardized factor score we call *teacher professionalism*, which was estimated from four principal-reported variables. Item descriptions and the properties of this factor score are provided in Appendix A, Table A2. The second school practice measure is a dummy-coded variable indicating whether SBM is practiced at the school. SBM is a school reform that involves decentralizing the decision-making authority from the district level to the school site, where it may involve teachers, parents, and other noncertificated school personnel, as well as the principal (David, 1989). There has been considerable debate about the impact of SBM on school effectiveness, particularly regarding whether SBM has a positive effect on student learning.

The lower panel of Table 7 shows the results of the group-level multinomial logistic regression part of the model, which uses practice variables to predict class membership. Class 1 is the reference category and coefficients are in the odds ratio metric. The intercept describes the odds of being a member of a given class when the predictors are zero, which corresponds to schools that do not use SBM and have an average level of Professionalism. The slope coefficients indicate the change in the odds of being in Classes 2 or 3 instead of Class 1 per unit change on the predictor concerned (slope coefficients less than 1.0 indicates a decrease in the odds and larger than 1.0 an increase). *Professionalism* was significantly associated with class membership. One standard deviation increase in the *Professionalism* score increases the odds of being in the very ineffective Class 2 instead of Class 1 ($\delta_{12} = .35, p < .01$) and the odds of being in the ineffective Class 3 instead of Class 1 by ($\delta_{13} = .51, p < .05$). In other words, the higher the *Professionalism*, the more likely that a school belongs to the effective Class 1. SBM, however, was not associated with class membership. These results are consistent with the literature on how these two measures of school practice are associated with school effectiveness.

Discussion

The MGMM is an important new analytic tool that can help researchers provide new perspectives on both new and old research problems. The usefulness of this model for addressing substantive research problems hinges upon on the presumption that the classes represent discrete subpopulations and not just

incidentally non-normality in the data (Bauer & Curran, 2004). That distinction can be difficult to verify or test statistically. Hence, for the applied researcher, the question of whether to pursue the MGMM should be based on whether adding classes can help advance knowledge and understanding of the given substantive problem. This is perhaps the essential, yet subjective, test of whether to categorize in any context, whether based on observed or unobserved classes.

Although the application presented in this article focuses on identifying substantively meaningful group-level classes that are the origin of non-normality in the random effects, the model is also appropriate for analyzing data with atheoretical non-normality (Ferguson, 1983; Vermunt & Van Dijk, 2001). That is, latent classes can be introduced to address violations in the normality assumption of unknown origin.

Much of the application we present focuses on model building decisions because those decisions can affect results. Of particular concern is the use of covariates to identify substantively meaningful classes. The addition of covariates can alter the distribution of the random effects and thus affect the number and composition of classes that are identified. To identify substantively meaningful classes, however, appropriate covariates should be determined *a priori* based on a conceptual framework informing the study. The arbitrary use of covariates in the classification model will likely result in arbitrary changes in the distributions of the random effects and these changes will undermine the identification of meaningful classes. This highlights the importance of the specification of covariates in the classification model and raises concerns about their use in an exploratory manner. In our example, we argued for a Type B school effects classification model that is conditional on student background and school social composition—factors that are known to affect student learning but are largely beyond the control of school site personnel. We do not include covariate controls in the classification model that measure school processes and practices because we wish to have the classes separated on those factors.

Whether to include random effects for the intercepts and slopes is another important consideration because the latent classes and random effects compete for the same variability in the trajectories. Recent research has shown that in some situations omitting the random effects can lead to overextraction (Bauer & Curran, 2004; Lubke & Neale, 2006). The results of this study seem to verify those findings. Compared with the unconditional classification model with the variance in the between-group intercepts and slopes constrained to zero, adding a random group-level intercept reduced the number of classes identified from five to three. Although more research is needed to better understand when omitting a random effect will lead to overextraction, the results of this study suggest that it is linked to whether covariates are included in the classification model. In our example, we saw that when the Type B control variables were in the classification model, retaining the random between-group intercept had no impact on the number and composition of the classes selected.

The MGMM is a new statistical tool with a wealth of applications in education and other social sciences, perhaps the most useful being research problems involving organizational effects where distinct subpopulations are hypothesized. The newness of this model offers considerable opportunity for addressing new research problems as well as readdressing old ones, while the complexity of the model results in additional specification considerations compared with traditional multilevel growth models.

Appendix A
Variable and Sample Descriptions

TABLE A1
Variable Descriptions and Early Childhood Longitudinal Study (ECLS) Labels

Variable Name	<i>M</i>	<i>SD</i>	Min	Max	Description, ECLS variable label, coding scheme
Measurement Variables (repeated within-student; Level 1, <i>N</i> = 13,604)					
Achievement Test Scores					
Math Fall K	20.47	7.39	7.08	56.91	Fall Kindergarten math IRT scale score (C1MRSCAL)
Math Spring K	28.81	8.79	8.34	59.75	Spring Kindergarten math IRT scale score (C2MRSCAL)
Math Fall 1st	33.75	9.31	8.10	60.03	Fall 1st grade math IRT scale score (C3MRSCAL)
Math Spring 1st	44.39	8.83	10.95	60.50	Spring 1st grade math IRT scale score (C4MRSCAL)
Time 1	2.18	0.51	1.15	4.31	Time in months from start of K to first data collection
Time 2	8.32	0.52	6.51	11.37	Time in months from start of K to second data collection
Time 3	13.26	0.67	11.80	17.13	Time in months from start of K to third data collection
Time 4	20.10	0.65	18.61	23.44	Time in months from start of K to fourth data collection
Summer Time	2.49	0.51	0.00	3.02	Time in months over the summer period

(continued)

TABLE A1 (continued)

Variable Name	<i>M</i>	<i>SD</i>	Min	Max	Description, ECLS variable label, coding scheme
Within-school (student) variables (Level 2, <i>N</i> = 3,401)					
SES	0.17	0.70	-0.99	2.88	Socioeconomic composite (WISESL)
Asian	0.04	0.20	0.00	1.00	(RACE = 5)
Black	0.14	0.34	0.00	1.00	(RACE = 2)
Hispanic	0.14	0.34	0.00	1.00	(RACE = 3 or 4)
Other	0.08	0.27	0.00	1.00	(RACE = 6, 7, or 8)
Between-school variables (Level 3, <i>N</i> = 2,54)					
Mean SES	0.14	0.49	-0.77	1.51	Classroom mean of (WISESL)
SB management	0.68	0.47	0	1	School-based management (S2MGTCMM)
Teacher professionalism	0.00	1.00	-3.59	1.00	Principal component measuring successful school emphasis on teacher professionalism (S2SUCC8-11)

TABLE A2
 Teacher Professionalism Factor Item Descriptions and Measurement Properties

Items Labels	Item Descriptions	Item Loadings
	Principal responses to, "How successful were your teachers at meeting these goals?"	
S2SUCC8	COMMUNICATES WELL WITH PARENTS	0.70
S2SUCC9	WORKING WELL WITH OTHER STAFF	0.76
S2SUCC10	OPEN TO NEW IDEAS/METHODS	0.80
S2SUCC11	PARTICIPATES IN PROFESSIONAL DEVELOPMENT	0.73
		Variance explained 56.2%

Note: All items are on a 3-point Likert-type scale (1 = not very successful, 3 = very successful).

Appendix B

Parameter Estimation by Maximum Likelihood

Based on the regression equations, the assumptions about the residuals in the repeated measurements, within-group, and between-group level submodels, and the hierarchical structure of the model, the density function for the response vector of an independent observation (in our case, a group or school) can be derived. We denote the probability density of the response vector \mathbf{Y}_j for group j conditional on the time variables, the within-group level covariates, and the between-group level covariates collected in the vectors \mathbf{a}_j , \mathbf{X}_j , and \mathbf{W}_j by $f(\mathbf{Y}_j|\mathbf{a}_j, \mathbf{X}_j, \mathbf{W}_j)$. The parameters of the MGMM can be estimated by means of maximum likelihood, which involves maximizing the sum of the log of the group-specific densities; that is, by maximizing:

$$\log L = \sum_{j=1}^N \log f(\mathbf{Y}_j|\mathbf{a}_j, \mathbf{X}_j, \mathbf{W}_j) \quad (\text{B1})$$

The density corresponding to a particular group can be constructed using the hierarchical structure of the multilevel mixture growth model. At the lowest level, we have the density corresponding to the response of subject i belonging to group j at time point t conditional on the unobserved random effects and the unobserved class memberships at the within- and between-group level and on the observed predictors at the various levels; that is, the conditional density $f(Y_{tij}|\mathbf{a}_{tij}, \mathbf{X}_{ij}, \mathbf{W}_j, \mathbf{u}_j, \mathbf{r}_{ij}, d_{ij} = 1, c_{ijk} = 1)$. This is a normal density with expected

(continued)

Appendix B (continued)

value $E(y_{tij}|\mathbf{a}_{tij}, \mathbf{X}_{ij}, \mathbf{W}_j, \mathbf{u}_j, \mathbf{r}_{ij}, d_{lj} = 1, c_{ijk} = 1)$ and variance σ^2 . Note that the expected value is a function of the unknown regression parameters and is obtained by substituting Equation 3a into Equation 2a and the resulting equation into Equation 1; that is, by writing the multilevel model in its mixed model formulation.

Using the density for the time-specific observations, we can construct the density at the within-group level conditional on the between-group level random effects, the class membership of school j , and the predictors in the model, which is denoted by $f(\mathbf{Y}_{ij}|\mathbf{a}_{ij}, \mathbf{X}_{ij}, \mathbf{W}_j, \mathbf{u}_j, d_{lj} = 1)$. This is achieved as follows:

$$\begin{aligned}
 & f(\mathbf{Y}_{ij}|\mathbf{a}_{ij}, \mathbf{X}_{ij}, \mathbf{W}_j, \mathbf{u}_j, d_{lj} = 1) \\
 &= \int \left(\sum_{k=1}^K \left[\prod_{t=1}^{T_{ij}} f(Y_{ijt}|\mathbf{a}_{tij}, \mathbf{X}_{ij}, \mathbf{W}_j, \mathbf{u}_j, \mathbf{r}_{ij}, d_{lj} = 1, c_{ijk} = 1) \right] P(c_{ijk} = 1) \right) f(\mathbf{r}_{ij}) d\mathbf{r}_{ij} \quad (\text{B2})
 \end{aligned}$$

As can be seen, the T_{ij} repeated measurements of subject i in group j are assumed to be independent observations conditionally on observed predictors, the random effects at the within and between level, and the class membership at both levels. The density for subject i is a marginal density that is obtained by summing over the K within-group classes and integrating over the within-group level random effects.

Finally, the marginal density for group j , $f(\mathbf{Y}_j|\mathbf{a}_j, \mathbf{X}_j, \mathbf{W}_j)$, is obtained as follows:

$$f(\mathbf{Y}_j|\mathbf{a}_j, \mathbf{X}_j, \mathbf{W}_j) = \int_{\mathbf{u}} \left(\sum_{l=1}^L \left[\prod_{i=1}^{n_j} f(\mathbf{Y}_{ij}|\mathbf{a}_{ij}, \mathbf{X}_{ij}, \mathbf{W}_j, \mathbf{u}_j, d_{lj} = 1) \right] P(d_{lj} = 1) \right) f(\mathbf{u}_j) d\mathbf{u}_j. \quad (\text{B3})$$

Here, the n_j observations of group j are assumed to be independent conditionally on observed predictors and between-group level random effects and class membership. The marginal density for group j is obtained by summing over the L between-group level classes and integrating over the between-group level random effects.

To summarize, maximum likelihood estimation involves maximizing the log likelihood function in Equation B1. The term appearing at the right-hand side of this equation is defined in Equation B3 and the densities appearing at the right-hand side of Equation B3 are defined in Equation B2.

For maximum likelihood estimation, one needs an algorithm and a method for solving the integrals appearing in Equations B2 and B3. As in standard multilevel analysis, the integrals appearing in Equation B2 can be solved analytically; that

(continued)

Appendix B (continued)

is, for each class separately and the results are summed up. The integrals in Equation B3 are solved by numerical integrations; that is, by Gauss-Hermite quadrature (Skrondal & Rabe-Hesketh, 2004). Latent GOLD uses a combination of the EM algorithm (Dempster, Laird, & Rubin, 1977) and Newton-Raphson to find the maximum of the log likelihood function. It begins with EM iterations and when close enough of the maximum switches to Newton-Raphson. Because the log likelihood of mixture models may contain local maxima, multiple random start values are used.

Model identification can be checked by determining whether the Jacobian matrix—containing the first-order derivatives of $f(\mathbf{Y}_j|\mathbf{a}_j, \mathbf{X}_j, \mathbf{W}_j)$ toward all model parameters—is full rank. This identification check can be requested in Latent GOLD.

Notes

1. We assume readers are familiar with multilevel growth models. Some comprehensive references on these subjects include the following texts: Bollen and Curran (2006); Duncan, Duncan, Strycker, Li, and Alpert (1999); Hedeker and Gibbons (2006); Raudenbush and Bryk (2002); Singer and Willett (2003); Skrondal and Rabe-Hesketh (2004).

2. This is not as critical a problem as it may seem. Indeed, the idea that any statistical model precisely reflects a real-life phenomenon is similarly difficult to establish. More important is whether the model is accurate enough to further our understanding of the phenomenon. See Cudeck and Henly (2003) for a detailed discussion on this issue.

3. Asparouhov and Muthen published two recent book chapters that have some overlap with the current study. These efforts were developed on parallel timelines as the current study and did not contribute to our ideas. Asparouhov and Muthen (2008) present an overview of multilevel mixture models with a focus on what they refer to as “within-between” models in which between-group classes are based on within-group classes. Muthen and Asparouhov (2009) provide a broad overview of growth mixture models and include a brief discussion of an MGMM with classes at the between-group level.

4. This article was first presented at the American Educational Research Association conference San Francisco on April 8, 2006, which is believed to be the first formal presentation of the MGMM.

5. Please see Palardy and Rumberger (2008) for a more detailed description of the ECLS data.

6. We provide an overview of the Type B school effect paradigm so that the application of the MGMM is grounded in a well-known example in the literature. We feel the usefulness of such demonstrations are enhanced when

grounded this way, particularly for highly complex models that require an array of specification decisions, because the substantive context of the analysis plays an important role in some of those decisions. However, we note that this is not a substantive investigation and we do not address the full complexity of the Type B effect model, which Raudenbush and Willms describe superbly.

7. Both SES measures are grand mean centered, which has implication on the interpretation of their effects on math achievement growth. When grand mean centered rather than group mean centered, individual SES accounts for the variation in the individual growth trajectory intercepts and slopes. Moreover, because the mean SES of the schools varies substantial across the sample, the grand mean centered individual SES variable also accounts for variance in the intercepts and slopes of the school trajectories. As a result, the coefficient on the mean SES is a compositional effect. That is, it accounts for variation in mean mathematics learning rates of the schools above and beyond the individual SES of the children.

8. Ideally, the fall tests would be administered at the very beginning of the school year and the spring test at the very end to maximize the precision of the learning estimates. However, the ECLS fall kindergarten tests were administered an average of 2.2 months after the start of the school year and ranged from 1.1 to 4.3 months after the start. The time between fall and spring tests also varied widely with a mean of 6.1 months and minimum and maximum values of 3.9 to 8.3 months. The first grade achievement test data collection schedule exhibited similar properties.

9. An alternative approach would be to select a classification model using likelihood-ratio tests comparing models differing in the numbers of classes with p values estimated by parametric bootstrap methods (McLachlan & Peel, 2000). We have, however, not encountered studies assessing the quality of this computationally intensive method for determining the number of between-group level classes in multilevel mixture models. This is one of the topics we leave for further research.

10. Although we propose using BIC-B in this situation, further research is needed on appropriate measures of information criterion for multilevel models in general and particularly multilevel mixture models where the classes are based on between-group level observations.

11. Note that we also estimated the MGMM with random effects for both the intercepts and the slopes, which also yielded a three-class model. However, that specification had higher values on the fit indices than the more parsimonious model with only random intercepts and therefore was rejected.

12. A likelihood ratio test could also be used to test for covariate invariance across the classes.

References

- Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55, 218–234.

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademia Kiado.
- Alexander, K. L., & Entwisle, D. R. (1996). Schools and children at risk. In Alan Booth & Judy Dunn (Eds.), *Family-school links: How do they affect educational outcomes?* (pp. 67–88). Mahwah, NJ: Lawrence Erlbaum.
- Asparouhov, T., & Muthen, B. (2008). Multilevel mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 27–75). Charlotte, NC: Information Age Publishing.
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8, 338–363.
- Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, 9, 3–29.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. New York, NY: John Wiley.
- Bozdogan, H. (1993). Choosing the number of component clusters in the mixture model using a new informational complexity criterion of the inverse-Fisher information matrix. In O. Opitz, B. Lausen, & R. Klar (Eds.), *Information and classification, concepts, methods and applications* (pp. 40–54). Berlin, Germany: Springer.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66, 227–268.
- Cudeck, R., & Henly, S. J. (2003). A realistic perspective on pattern representation in growth data: Comment on Bauer and Curran. *Psychological Methods*, 8, 378–383.
- David, J. L. (1989). Synthesis of research on school-based management. *Educational Leadership*, 46, 45–53.
- Day, N. E. (1969). Estimating the components of mixture of normal distributions. *Biometrika*, 56, 463–474.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B*, 39, 1–38.
- Dias, J. G. (2006). Model selection for the binary latent class model: A Monte Carlo simulation. In V. Batagelj, H.-H. Bock, A. Ferligoj, & A. Žiberna (Eds.), *Data science and classification* (pp. 91–99). Berlin, Germany: Springer.
- Dias, J. G. (2007). Model selection criteria for model-based clustering of categorical time series data: A Monte Carlo study. In R. Decker & H.-J. Lenz (Eds.), *Advances in data analysis* (pp. 23–30). Berlin, Germany: Springer.
- Duncan, T. E., Duncan, S. C., Strycker, A. L., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Ferguson, T. S. (1983). Bayesian density estimation via mixtures of normal distributions. In M. H. Rizvi, J. S. Rustagi, & D. Siegmund (Eds.), *Recent advances in statistics* (pp. 287–302). New York, NY: Academic Press.
- Fox, J. P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, 66, 269–286.

- Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. *American Journal of Sociology*, 79, 1179–1259.
- Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: John Wiley.
- Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, E. A. Suchman, P. F. Lazarsfeld, S. A. Starr, & J. Clausen (Eds.), *Studies on social psychology in World War II, Vol. 4: Measurement and prediction*. Princeton, NJ: Princeton University Press.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.
- Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, 41, 499–532.
- Lukociene, O., & Vermunt, J. K. (2010). Determining the number of components in mixture models for hierarchical data. In A. Fink, L. Berthold, W. Seidel, & A. Ultsch (Eds.), *Advances in data analysis, data handling and business intelligence*. Berlin, Germany: Springer.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: John Wiley.
- Muthen, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: APA.
- Muthen, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Thousand Oaks, CA: Sage.
- Muthen, B., & Asparouhov, T. (2009). Growth mixture analysis: Models with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Advances in longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC Press.
- Muthen, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463–469.
- Muthen, L. K., & Muthen, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Author.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, 4, 139–177.
- Nagin, D. S. (2005). *Group-based modeling of development*. Boston, MA: Harvard University Press.
- Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods*, 6, 18–34.
- National Center for Education Statistics (NCES). (2002). User's manual for the ECLS-K first grade public-use data files and electronic codebook. NCES 2002–135.
- Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in the first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, 30, 111–140.
- Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata* (2nd ed.). College Station, TX: Stata Press.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69, 167–190.

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307–335.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal and structural equation models*. London, England: Chapman & Hall/CRC.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *Winbugs user manual version 1.4*. Retrieved August 3, 2006, from <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>
- Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelson (Eds.), *Advances in latent variable mixture models*. Charlotte, NC: Information Age Publishing.
- Verbeke, G., & Lesaffre, G. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91, 217–221.
- Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, 33, 213–239.
- Vermunt, J. K., & Magidson, J. (2005). *LatentGOLD 4.0 user's manual*. Belmont, MA: Statistical Innovations.
- Vermunt, J. K., & Magidson, J. (2008). *LG-syntax user's guide: Manual for Latent GOLD 4.5 syntax module*. Belmont, MA: Statistical Innovations.
- Vermunt, J. K., & Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modeling Newsletter*, 13, 6–13.
- Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5, 329–350.

Authors

GREGORY PALARDY is an Assistant Professor in the Graduate School of Education at the University of California, 2105 Sproul Hall, 900 University Ave., Riverside, CA 92507; e-mail: gjpalardy@gmail.com. His research interests are on multilevel latent variable models including longitudinal, mixture, and crossed random effect models.

JEROEN VERMUNT is a full professor in the Department of Methodology and Statistics at Tilburg University, PO Box 90153, 5000LE Tilburg, Netherlands; e-mail: j.k.vermunt@uvt.nl. His research is on methodology of social, behavioral, and biomedical research, with a special focus on latent variable models and techniques for the analysis of categorical, multilevel, and longitudinal data. He has widely published on these topics in statistical and methodological journals and has co-authored many articles in applied journals in which these methods are used to solve practical research problems. He is the co-developer (with Jay Magidson) of the Latent GOLD software package. In 2005, Vermunt received the Leo Goodman award from the Methodology Section of the American Sociological Association.

Manuscript received September 19, 2008

Revision received June 5, 2009

Accepted August 18, 2009