The Impact of Ignoring Cross-Classified Multiple Membership Data Structures

National Council on Measurement in Education
Chicago, IL

Daniel L. Murphy
Burcu Kaniskan
Ahmet Turhan

April 2015

ABSTRACT

This study compared the use of a three-level growth-curve model with that of a cross-classified growth curve model and a cross-classified multiple membership growth-curve model for handling cross-classified multiple membership data structures. Results indicate that growth models which ignore the complexity of cross-classified multiple membership data structures produce negatively biased estimates of teacher random effects variance components. We recommend that teacher evaluation systems use growth models that can accommodate the complexity of cross-classified multiple membership longitudinal data structures when using student growth data to estimate teacher effectiveness.

During the recent decade, policies for teacher evaluation in U. S. public K-12 education have changed dramatically. Perhaps the most significant policy changes incorporate the use of student growth data to evaluate teacher effectiveness. For example, in 2009, President Obama announced that U.S. Department of Education (USED) would provide funding for a competitive grant program called Race to the Top to generate innovations and reforms in state and local public K-12 education. To qualify for grant funding, applicants were required to develop an evaluation system that provides annual teacher effectiveness ratings based in part on student achievement growth.

In addition, in 2011, the USED invited each state educational agency to request flexibility from specific requirements of the Elementary and Secondary Education Act (ESEA), also known as No Child Left Behind (NCLB, 2001), in exchange for state-developed plans designed to improve educational outcomes for all students (Dunlap, 2011). To request ESEA flexibility, state education agencies were required to develop teacher and principal evaluation systems that provide effectiveness ratings using student achievement growth data as a significant factor. In response to such policy initiatives, many states are either developing or assessing methodologies that can be used to analyze student achievement growth to evaluate teacher effectiveness.

Measuring student growth requires the collection of student data on multiple occasions and the application of a statistical model to describe student progress over time. When using these statistical models to evaluate teacher effectiveness, it is important to examine the underlying assumptions that accompany them. For example, multilevel modeling techniques enable the measurement of student growth by considering repeated measures of student achievement (e.g., standardized test scores) as nested within students. If students are considered to be nested within teachers, then teacher effectiveness can be incorporated into the model.

Under this approach, growth is modeled as varying across students at Level-2 and teachers at Level-3, enabling teachers' effects on student growth to be estimated (Grady & Beretvas, 2010). The use of such a model assumes, however, that students remain perfectly nested within the same teachers for each of the measurement occasions.

Unfortunately, longitudinal data sets with educational data in practice typically do not meet this assumption, presenting complications that require use of more complex multilevel modeling techniques. For example, when measuring student growth across grade levels, if students change teachers from one measurement occasion to another, the fundamental assumption that lower level units (students) are purely nested in higher level units (teachers) is violated. In this case the data are considered to have a *cross-classified* rather than a purely nested data structure. Previous research indicates that modeling cross-classified data as purely nested can result in biased estimates of the random effects variances and standard errors of interest (Luo & Kwok, 2009, 2012; Meyers and Beretvas, 2006).

Moreover, there is a further complication that occurs when students encounter multiple teachers within the same school year who contribute to a single test score. If students have more than one teacher in the same subject area, or change classes within a school year, the data are considered to have a *multiple membership* structure. It is well documented that ignoring multiple membership data structures also can result in biased estimates of the random effects variance components and standard errors of interest (Chung & Beretvas, 2011; Grady & Beretvas, 2010; Goldstein, Burgess, & McConnell, 2007; Leckie, 2009).

The body of research documenting the biased estimates that occur when multilevel growth model assumptions are violated has important implications for teacher evaluation systems that use student achievement growth data. In particular, the research suggests that estimates of teacher effects can be biased if the models do not sufficiently account for the complexity of the data structure. If the teacher effects estimates are biased, then the likelihood that teachers will be misclassified within the evaluation system increases. There can be high stakes ramifications associated with such misclassifications; therefore, correctly modeling the complexity of the data structure is critical.

Real world longitudinal data sets containing educational data not only contain cross-classified and multiple membership data structures, they often combine both structures within the same data set. To the best of our knowledge, the effects of ignoring cross-classified multiple membership data structure estimates when modeling student achievement growth has not yet been studied. The purpose of this paper therefore is to initiate such an investigation.

To investigate the effects of mismodeling cross-classified multiple membership data, we introduce a cross-classified multiple membership growth curve model (*CCMM-GCM*) to properly account for the complexity of the data structure that extends the cross-classified growth curve model (*CC-GCM*) developed by Raudenbush and Bryk (2002). We then compare three methods for handling cross-classified multiple membership data structures: 1) a three-level multilevel growth curve model (*GCM*) that ignores the cross-classified and multiple membership data structures, 2) a *CC-GCM* that ignores only the multiple membership data structure, and 3) a *CCMM-GCM* that models the data structure appropriately. The next section of the paper describes these three methods in detail.

Background

CGM

Bryk and Raudenbush (1988) provide a framework for modeling student growth in educational contexts where students are nested within teachers. Under this framework, repeated measures of student outcomes (i.e., Level-1) are viewed as nested in students (Level-2), and students are viewed as nested within teachers (Level-3). Using Raudenbush and Bryk's (2002) formulation, Level-1 of a three-level *GCM* would be,

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}a_{tij} + e_{tij}, \tag{1}$$

where Y_{tij} is the outcome score at time *t* for student *i* with teacher *j*, π_{0ij} is the score for student *i* with teacher *j* at the time point coded as zero (often the initial measurement), π_{1ij} is the linear growth rate (per academic year in this study) for student *i* with teacher *j*, and e_{tij} is the Level-1 residual. The Level-1 residuals are assumed to be normally distributed with a mean of zero (i.e., $e_{tij} = N(0, \sigma^2)$).

At Level-2, the equations would be,

$$\begin{cases} \pi_{0ij} = \beta_{00j} + r_{0ij} \\ \pi_{1ij} = \beta_{10j} + r_{1ij} \end{cases}, \tag{2}$$

where β_{00j} is the mean score for students with teacher *j* at the time point coded as zero, r_{0ij} is the random effect for student *i* with teacher *j* at time zero, β_{10j} is the mean yearly growth rate for teacher *j*, and r_{1ij} is the deviation of the yearly growth rate for student *i* from the mean growth rate for teacher *j*. The Level-2 random effects are assumed to be normally distributed with means of zero and the following covariance structure: $cov\begin{bmatrix} r_{0ij} \\ r_{1ij} \end{bmatrix} = \begin{bmatrix} \tau_{r00} & \tau_{r01} \\ \tau_{r10} & \tau_{r11} \end{bmatrix}$.

At Level-3, the equations would be,

$$\begin{cases} \beta_{00j} = \gamma_{000} + u_{00j} \\ \beta_{10j} = \gamma_{100} + u_{10j} \end{cases}, \tag{3}$$

where γ_{000} is the grand mean of the outcome across teachers at time zero, u_{00j} is the random effect for teacher *j*, γ_{100} is the grand mean growth rate across teachers, u_{10j} is the random effect of teacher *j* on the students' mean growth rate. The Level-3 random effects also are assumed to be normally distributed with means of zero, and $cov\begin{bmatrix} u_{00j} \\ u_{10j} \end{bmatrix} = \begin{bmatrix} \tau_{u00} & \tau_{u01} \\ \tau_{u10} & \tau_{u11} \end{bmatrix}$.

When the data are purely clustered and the initial time point is coded as zero, the random intercepts under the *GCM* reflect the fact that students and teachers do not begin with the same level of achievement. The random growth rates or slopes at Level-3 reflect different growth rates associated with each teacher. Thus, teachers responsible for well above average yearly growth rates could be considered to be highly effective.

An alternative specification for the model in Equation 3 results when the growth slope at Level-3 is considered to be fixed. This model fits the data when the variation of yearly growth rate across teachers is considered to be insignificant. In that case the Level-3 equations become,

$$\begin{cases} \beta_{00j} = \gamma_{000} + u_{00j} \\ \beta_{10j} = \gamma_{100} \end{cases}, \tag{4}$$

and the covariance structure is reduced to $u_{00j} = N(0, \tau_{u00})$. Under this model, positive teacher effects would add to students' growth, and negative teacher effects would offset students' growth. Figure 1 illustrates the concept behind this configuration of the GCM.

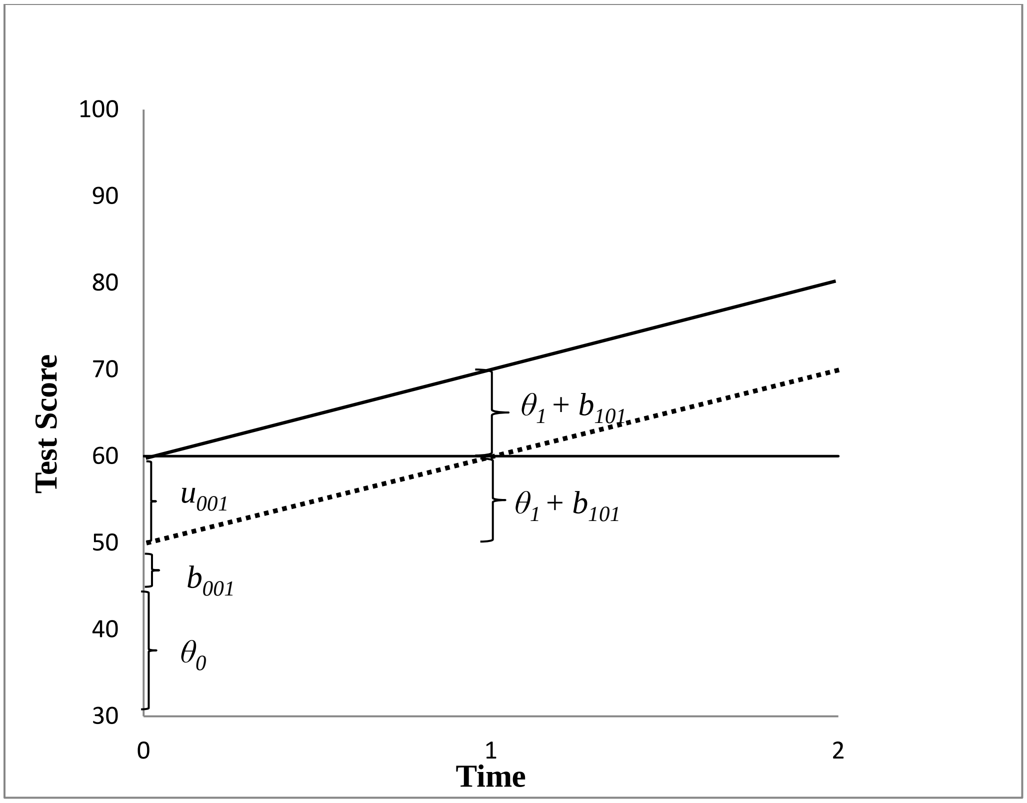

Figure 1. Growth trajectory (dotted line) and expected deflection (solid line) for student b_{001} who encounters teacher u_{001}.

Although the *GCM* is appropriate when data are purely clustered, educational data sets in practice typically contain complications that require use of more complex multilevel modeling techniques. It is typical, for example, for students to change classrooms and teachers from one year to the next. Longitudinal data sets containing such complicated student data present a methodological challenge for researchers who use *GCM*s to model growth. Because the assumption that lower level units are purely clustered in higher level units is violated when students encounter multiple teachers across time, the *GCM* would not fit the data. Under this scenario, the data are considered to have a cross-classified data structure.

Cross-classified data can be misspecified by either omitting or ignoring the crossed factor. A crossed factor would be omitted, for example, when students are cross-classified by schools and neighborhoods and the effects on student achievement of living in a certain neighborhood are omitted from the model (see Luo and Kwok, 2009; Meyers and Beretvas, 2006). A crossed-factor would be ignored, for example, when a *GCM* is specified to analyze school level growth, but some students move from one year to the next and change schools. In that case, specifying a purely nested relationship oversimplifies the data structure by ignoring the crossed classification caused by students transferring from one school to another across time (Luo & Kwok, 2012). Previous research indicates that misspecifying cross-classified data by either omitting or ignoring the crossed factors can result in biased estimates of the random effects variances and standard errors of interest (Luo & Kwok, 2009, 2012; Meyers and Beretvas, 2006).

This body of research is important, because it suggests that growth (i.e., value-added) models relying on random effects estimates may produce biased teacher effectiveness results if cross-classified data structures are omitted or ignored. Fortunately, *GCM*s have been extended, and *CC-GCM*s have been developed to fit cross-classified data structures (Goldstein, 2010; Rasbash & Goldstein, 1994; Raudenbush, 1993, Raudenbush & Bryk, 2002). A two-level CC-GCM which can be used to estimate student and teacher random effects is described in the following section.

CC-GCM

Raudenbush and Bryk (2002) introduced the following two-level *CC-GCM*, which allows modeling of student and teacher effects when students change teachers from one measurement occasion to another. The Level-1 model under this formulation is the same as the Level-1 model under the *GCM* described in Equation 1. In contrast with the Level-2 model under the *GCM* described in Equation 2, the *CC-GCM* model would be,

(5)

$$\begin{cases} \pi_{0ij} = \theta_0 + b_{00i} + c_{00j} \\ \pi_{1ij} = \theta_1 + b_{10i} \end{cases},$$

where θ_0 is the mean of the outcome at time zero, b_{00i} is the random effect for student *i* at time zero, θ_1 is the mean yearly growth rate, b_{10i} is the deviation of the yearly growth rate of student *i* from the mean growth rate, and c_{00j} is the random effect associated with teacher *j*. The random effect c_{00j} is interpreted as an expected deflection to a student's growth rate resulting from encountering teacher *j* at time *t* (Raudenbush and Bryk, 2002).

In the model described in Equation 5, the teacher effect disappears after the end of each year. Hence, this model is sometimes referred to as a "non-persistent model." To incorporate persistent (sometimes referred to as cumulative) teacher effects, Raudenbush and Bryk (2002) use a dummy variable at Level-2 (D_{tij}), which allows the teacher effects to accumulate over time. In this "persistent" model the Level-2 equations become:

$$\begin{cases} \pi_{0ij} = \theta_0 + b_{00i} + \sum_{j=1}^{J} \sum_{t=0}^{T} D_{tij}\, c_{00j} \\ \pi_{1ij} = \theta_1 + b_{10i} \end{cases}, \tag{6}$$

where the dummy variable $D_{tij} = 1$ if student *i* has encountered teacher *j* at or prior to time *t*, and is 0 otherwise. The double summation in Equation 5 accumulates the effects of teacher c_{00j} across time (Raudenbush & Bryk, 2002). The Level-2 residuals of the *CC-GCM* are assumed to have the following covariance structure: $\begin{pmatrix} b_{00i} \\ b_{10i} \end{pmatrix} \sim N\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \tau_{b00} & \tau_{b00,10} \\ \tau_{b10,00} & \tau_{b10} \end{bmatrix}, c_{00j} \sim (0, \tau_{c00})$.

To demonstrate the formulation of the persistent cross-classified model, consider a hypothetical case where student *i* encounters teacher 1 during year one, teacher 2 during year 2, and teacher 3 during year three. The predicted score for student *i* at each of the three years would be:

Predicted value at year one: $Y_{1i1} = \theta_0 + b_{00i} + c_{001}$

Predicted value at year two: $Y_{2i2} = \theta_0 + b_{00i} + \theta_1 + b_{10i} + c_{001} + c_{002}$

Predicted value at year three: $Y_{3i3} = \theta_0 + b_{00i} + 2(\theta_1 + b_{10i}) + c_{001} + c_{002} + c_{003}$

As demonstrated, the students' gains each year are a function of their growth rate and teachers' effects that persist and accumulate in the model. This formulation of the persistent *CC-GCM* is illustrated in Figure 2.

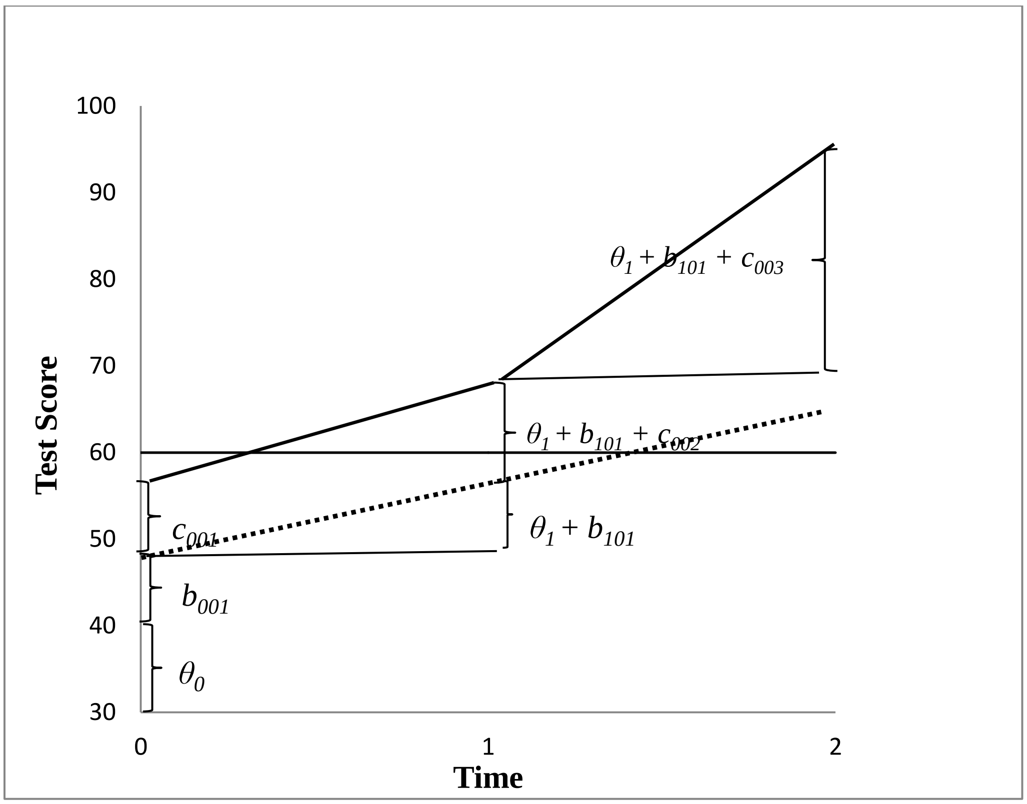

Figure 2. Growth trajectory (dotted line) and expected deflections that accumulate (solid lines) for student b_{001} who encounters teachers c_{001}, c_{002}, and c_{003} under a cross classified random effects growth model.

While *CC-GCMs* can handle cross-classified data structures resulting from students changing teachers between measurement occasions, there is a further complication that occurs when students encounter multiple teachers who contribute to a single test score within the same school year. For example, it is not uncommon for secondary school students to receive a schedule change part way through the school year (e.g., to add an honors class, drop an honors class, or change an elective class). Those schedule changes often result in students changing teachers, potentially resulting in more than one teacher having an effect on a summative test score.

Another example of multiple teachers contributing to a single test score occurs when students have supplementary classes in a given subject area for enrichment or remediation purposes. It is a common practice, for example, to assign students who perform poorly on a summative exam to a supplemental class. These remedial classes are designed not to replace but to support the students' more general subject-area class. In this instance, the students have (at least) two teachers who contribute to learning that is measured on a subsequent summative exam. When more than one teacher contributes to the same student achievement measure, the data are considered to have a multiple membership data structure. Previous research indicates that ignoring multiple membership data structures can result in biased estimates of the random effects variance components and standard errors of interest (Chung & Beretvas, 2011; Grady & Beretvas, 2010; Goldstein, Burgess, & McConnell, 2007; Leckie, 2009).

To summarize, as discussed above, failure to model cross-classified data structures can have adverse effects on the accuracy of parameter estimates (e.g., based estimates of random effects variance components). Similarly, failure to model multiple membership data structures can have adverse effects on the accuracy of parameter estimates. To the best of our knowledge, the effects of ignoring a longitudinal data structure composed of cross-classified factors and multiple membership structures have not yet been studied, though such structures are likely to be seen in practice. Fortunately, random effects models have been extended to accommodate such complex data. The next section of the paper introduces a growth model that measures random student and teacher effects and can accommodate both cross-classified and multiple membership data structures.

CCMM-GCM

The fact that students may have more than one teacher contributing effects to a single test score presents a methodological dilemma for researchers interested in modeling the causal effects of teachers on student achievement. Multiple membership data structures also violate the assumption that lower level units are perfectly nested in higher level units, so *GCM*s do not fit the data. While *CC-GCM*s account for students being cross-classified by teachers across time, each measurement occasion under the *CC-GCM* is linked to one teacher. When the data possesses a multiple membership structure, properly specified models must account for a set of teachers contributing to the outcome from a single measurement occasion. The typical method for doing so assigns weights to the teachers.

There is more than one way to assign weights in a multiple membership scenario. One method assigns equal weights to each teacher encountered regardless of the proportion of time that a student encountered each teacher. Another method assigns weights that reflect the length of time a student encountered each teacher (i.e., dosage), potentially resulting in unequal weights. In the current study the first strategy was employed. In particular, if a student was a member of only one classroom in a given year, then a weight of 1 was associated with the relevant teacher. For students encountering two teachers within a given year, each teacher was assigned a weight of 1/2. For students encountering three teachers within a given year, each teacher was assigned a weight of 1/3.

To accurately model the case where students are cross-classified across school years while having a multiple membership data structure within school years, the cross-classified model described in Equations 5 and 6 must be extended to account for the multiple membership structure. This is done by linking sets of teachers to each outcome. The notation for the *CCMM-GCM* at Level 1 therefore would be,

$$Y_{ti\{k\}} = \pi_{0i\{k\}} + \pi_{1i\{k\}} a_{ti} + e_{ti\{k\}}, \tag{7}$$

where $Y_{ti\{k\}}$ represents the outcome at time t for student i taught by the set of teachers $\{k\}$, and $e_{ti\{k\}}$ represents the residual at time t for student i given the set of teachers $\{k\}$. The Level-2 model would be,

$$\begin{cases} \pi_{0i\{k\}} = \theta_0 + b_{00i} + \sum_{h \in \{k\}} \sum_{t=0}^{T} D_{ti\{k\}} w_{tih} c_{00h} \\ \pi_{1i\{k\}} = \theta_1 + b_{10i} \end{cases} \tag{8}$$

In this model, $D_{ti\{k\}} = 1$ if student i has encountered the set of teachers $\{k\}$ at or previous to time t, and is 0 otherwise. The term b_{00i} is the random effect for student i at the time point coded zero after controlling for the weighted average of the effects of the set of $\{k\}$ teachers. The term w_{tih} represents the weight at time t for student i associated with teacher $h \in \{k\}$. The coefficient, b_{10i}, is the deviation of the growth rate of student i from the average growth rate, as described earlier. The Level-2 random effects are assumed to have the following covariance structure:

$$\begin{pmatrix} b_{00i} \\ b_{10i} \end{pmatrix} \sim N\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \tau_{b00} & \tau_{b00,10} \\ \tau_{b10,00} & \tau_{b10} \end{bmatrix}, c_{00\{k\}} \sim (0, \tau_{c00}).$$

In the context of modeling teachers' effects on student growth, failure to correctly link students who transfer classrooms, are team taught, or have other complex links with teachers will produce bias in the model results (McCaffrey, Lockwood, Koretz, & Hamilton, 2004). To the best of our knowledge, the effects of ignoring a longitudinal data structure composed of cross-classified factors and multiple membership structures has not yet been studied, though such structures are likely to be seen in practice. Therefore, the purpose of this study is to examine the nature of bias introduced when growth models do not sufficiently account for the complexity of cross-classified multiple membership data structures.

Method

This study used simulated data to compare three models used for handling cross-classified multiple membership data structures. The three compared models were: 1) a traditional *GCM* that ignored both the cross-classified and multiple membership data structures, 2) a *CC-GCM* that ignored only the multiple membership data structure, and 3) a *CCMM-GCM* that modeled the cross-classified multiple membership data structure appropriately.

Conditions

The experimental design for this study was a 2 x 2 x 3 factorial analysis, resulting in 12 examined conditions. The manipulated conditions were: the percentage of students who encountered multiple teachers within the same school year (15%, 30%), the number of teachers encountered by those students (2, 3), and the estimating model used to analyze the data (*GCM*, *CC-GCM*, *CCMM-GCM*). SAS software version 9.2 (SAS Institute Inc, Cary, North Carolina) was used to generate 500 data sets per condition for a total of 6000 data sets.

The percentages of students with multiple membership data structures were selected to mimic a sample of middle school (i.e., grades 6-8) data from a large urban district. Examination of that sample data set indicated that the percentage of students encountering multiple teachers within a single year ranged from approximately 20% to 30% over three school years. The values of 15% and 30% were selected based on the real data set to mimic scenarios where small and moderate percentages of students respectively belong to multiple classrooms.

The values for the numbers of teachers encountered by multiple membership students in this study were selected to facilitate interpretation and are based on Chung & Beretvas's (2011) research examining the impact of ignoring multiple membership data. It is important to note that multiple the membership structures seen within the real data set were considerably more complex than the structures examined in this study.

Data Generation

To mimic findings from the real data set, 3,347 students with measurements at each of three time points were included in the base data set. The number of teachers per grade ranged from 79 in the first year to 94 in the third year. The average number of students per teacher was 34, ranging from 7 to 127. The model used for data generation was the *CCMM-GCM* described earlier in the paper in Equations 7 and 8.

The data were generated to include an initial measurement and the effects of time using the Level-1 equation described in Equation 7, where $Y_{ti\{k\}}$ represents the outcome at time *t* for student *i* taught by the set of teachers $\{k\}$. The residual at time *t* of the score from student *i* from the expected score given the set of teachers $\{k\}$ was generated as $e_{ij\{k\}} \sim N(0, 0.40)$. The generating values for the fixed effects coefficients were 0.10 for the intercept (θ_0), and 0.50 for the growth rate (θ_1).

The Level-2 equations described in Equation 8 were used with the following random effect variance parameters: $\left\{\begin{pmatrix} b_{00i} \\ b_{10i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.20 & 0.05 \\ 0.05 & 0.10 \end{pmatrix}\right], c_{00\{k\}} \sim N(0, 0.20)\right.$. These values were selected based on values used in previous cross-classified growth model simulation studies

(Luo & Kwok, 2012). The intraclass correlation, defined as $ICC = \frac{\tau_{c00}}{\tau_{c00}+\tau_{b00}+\sigma^2}$, was therefore set at 25%, which is within the range of values commonly seen in previous methodological and applied research (e.g., Chung & Beretvas, 2011; Hox, 2002; Luo & Kwok, 2012).

To generate multiple membership patterns, the random teacher effects, $c_{00\{k\}}$, were weighted and summed so that multiple membership students received a weighted average teacher effect from the set of teachers $\{k\}$. The coefficient w_{tih} in Equation 8 represents the weight at time *t* for student *i* associated with teacher $h \in \{k\}$. The weights were assigned equally to each teacher associated with the student and sum to one, that is, $\sum_{h\in\{k\}} w_{tih} = 1$. If, for example, a student was a member of three different classrooms at a given time point, then each teacher was assigned a weight of 1/3. If a student was a member of only one classroom at a given time point, then a weight of 1 was associated with the relevant teacher.

Data Analysis

Markov chain Monte Carlo (MCMC) as implemented in MLwiN software (version 2.27, Rasbash, Charlton, Browne, Healy, & Cameron, 2009) was used to estimate each model, including the use of the default priors for the variance distributions. To facilitate software convergence, vaguely informative starting values of that varied between 0.10 and 0.30 based on a sample size of 5 were implemented for each random teacher effect variance parameter. Using the small sample size of 5 reduced the influence of the starting values on the final parameter estimates, so that model estimates were free to override the starting values based on the data. Each chain was run with 50,000 iterations with a burn-in of 5,000.

Three different models were fitted to each cross-classified multiple membership data set that was generated. The *GCM* described in Equations 1, 2 and 4 that ignores cross-classification and multiple membership, the *CC-GCM* described in Equations 1 and 6 that ignores multiple membership, and the correct *CCMM-GCM* described in Equations 7 and 8.

Parameter estimates were saved for each of the 500 replications under each of the models. The parameters' recovery was evaluated using the relative parameter bias given by

$$RB(\omega) = \frac{\bar{\hat{\omega}} - \omega}{\omega}$$

where $\bar{\hat{\omega}}$ is the mean parameter estimate, and ω is the true parameter value. Estimates were considered biased when the absolute value of relative bias exceeded .05 (Hoogland & Boomsma, 1998).

Following the recommendation of Hauck and Anderson (1984) that simulation studies be analyzed using the same tools as other experimental studies, analyses of variance (ANOVAs) were used to examine which simulation condition factors affected the relative bias of parameter estimates. Because of the large number of observations, standard probability values would not necessarily indicate that a factor had a practically significant effect. Therefore, only factors that were statistically significant (p < .01) predictors of relative bias and that were associated with effect size measure values (η^2) greater than .01 were considered to be practically significant. Previous multilevel model simulation studies (Chung & Beretvas, 2011; Krull & MacKinnon, 1999; Murphy & Pituch, 2009) have used similar criteria to identify factors affecting parameter bias.

Results

Convergence

The proportions of non-convergent solutions across conditions are presented in Table 1. All of the models produced non-convergent solutions across each condition. The non-convergent solutions occurred because of negative estimates for at least one of the Level-2 variance parameters. In general, all models produced more non-convergent solutions when 15% of the students had multiple membership structures than they did when 30% of the students had such structures. In addition, under the conditions where multiple membership students encountered two teachers, the correctly specified *CCMM-GCM* produced more non-convergent solutions than did the misspecified *GCM* and *CC-GCM*.

Table 1

Proportion of Non-convergent Solutions for the GCM, CC-GCM, and CCMM-GCM across Multiple Membership Student Percentage (mm%) and Number of Teachers Encountered (n) Conditions

		Estimating Model		
mm%	*n*	*GCM*	*CC-GCM*	*CCMM-GCM*
15%	2	.084	.078	.100
	3	.058	.056	.056
30%	2	.048	.046	.062
	3	.054	.054	.054

Fixed Effects Estimates

No substantial relative bias was detected for the intercept (θ_0) or growth slope (θ_1) fixed effect point estimates across the examined conditions, with one exception: the intercept point estimate under the correctly specified *CCMM-GCM* was slightly outside the acceptable range. Relative bias means and standard deviations for estimates of the θ_0 under each model were: M = 0.005 and SD = 0.585 for *GCM* estimates, M = -0.029 and SD = 0.534 for *CC-GCM* estimates, and *M* = 0.057 and *SD* = 0.535 for *CCMM-GCM* estimates. Relative bias means and standard deviations for estimates of θ_1 under each model were: *M* < 0.001 and *SD* = 0.081 for *GCM* estimates, *M* = −0.001 and *SD* = 0.075 for *CC-GCM* estimates, and *M* = 0.001 and *SD* = 0.074 for *CCMM-GCM* estimates.

Random Effects Variance Component Estimates

Level-1 error variance

The *CC-GCM* and *CCMM-GCM* produced estimates of σ^2 that were within the acceptable range of relative bias, as shown in Table 2.

Table 2

Summary of Mean Relative Parameter Bias by Model for Parameter Estimates of the Level-1 Error Variance component, σ^2*, Level-2 Random Student Intercept Variance Component,* τ_{b00}*, Level-2 Covariance between the Intercept and Slope,* $\tau_{b00,10}$*, Level-2 Random Student Slope Variance Component,* τ_{b10}*, Level-2 Random Year-one Teacher Effect Variance Component,* $\tau_{c00(1)}$*, Level-2 Random Year-two Teacher Effect Variance Component,* $\tau_{c00(2)}$*, and Level-2 Random Year-three Teacher Effect Variance Component,* $\tau_{c00(3)}$

	Parameter						
Model	σ^2	τ_{b00}	$\tau_{b00,10}$	τ_{b10}	$\tau_{c00(1)}$	$\tau_{c00(2)}$	$\tau_{c00(3)}$
GCM	0.146	-0.118	0.596	0.554	-0.205	—	—
	(0.036)	(0.213)	(0.330)	(0.158)	(0.238)		
CC-GCM	0.022	0.073	0.104	0.053	-0.221	-0.231	-0.241
	(0.028)	(0.092)	(0.202)	(0.101)	(0.168)	(0.161)	(0.153)
CCMM-GCM	0.008	0.005	0.038	0.013	0.032	-0.022	-0.021
	(0.028)	(0.091)	(0.201)	(0.102)	(0.184)	(0.176)	(0.172)

Note. — = not applicable.

By contrast, the *GCM*, which ignores both cross-classified and multiple membership data structures, produced estimates with significant positive relative bias. Given the identification of substantial bias under the *GCM*, an ANOVA was conducted to examine the factors associated with the relative bias. As shown in Table 3, ANOVA results revealed that the relative bias of Level-1 error variance estimates depended on the percentage of multiple membership students in the data and the number of teachers encountered by those students.

Table 3

Analysis of Variance Results for Mean Relative Bias of the Level-1 Error Variance Component, σ^2, Level-2 Random Intercept Variance Component, τ_{b00}, Level-2 Covariance between the Intercept and Slope, $\tau_{b00,10}$, and Level-2 Random Slope Variance Component, τ_{b10}

	σ^2		τ_{b00}		$\tau_{b00,10}$		τ_{b10}	
Source	*F-ratio*	$\eta^2_{partial}$	*F-ratio*	$\eta^2_{partial}$	*F-ratio*	$\eta^2_{partial}$	*F-ratio*	$\eta^2_{partial}$
				Between subjects				
mm% (P)	81.57*	**.039**	55.22*	**.026**	0.08	< .001	17.08*	.008
t (T)	25.17	**.012**	6.25	.003	0.02	< .001	0.68	< .001
P x T	0.19	< .001	1.40	< .001	0.58	< .001	0.05	< .001
				Within subjects				
Model (M)	—	—	1676.11*	**.456**	5683.94*	**.740**	29593.50*	**.937**
M x P	—	—	0.64	< .001	51.97*	**.025**	302.44*	**.132**
M x T	—	—	1.22	< .001	8.41*	.004	29.47*	**.015**

Note. — = not applicable. Practically significant $\eta^2_{partial}$ values ($\eta^2_{partial}$ > . 01) appear in bold face. *p < .01.

The magnitude of positive relative bias produced by the *GCM* decreased from 0.153 to 0.139 as the percentage of multiple membership students in the data increased from 15% to 30%. Similarly, the magnitude of positive relative bias decreased from 0.150 to 0.142 as the number of teachers encountered by multiple membership students increased from two to three. This finding suggests that a data set containing small percentages of students with cross-classified multiple membership data structures can result in estimates of σ^2 with severely positive bias when such complex structures are ignored under the *GCM*. Somewhat surprisingly, the positive bias was less severe when more students had multiple membership data structures, and those students encountered more teachers.

Student random effects variance components

As shown in Table 2, *CCMM-GCM* estimates of the student random intercept variance (τ_{b00}), random slope variance (τ_{b10}), and covariance between the intercept and slope ($\tau_{b00,10}$) were not substantially biased. By contrast, inappropriate modelling of the cross-classified and multiple membership structures under the *GCM* and *CC-GCM* resulted in biased estimates of all three student random effects variance components.

Given the identification of substantial bias for estimates of τ_{b00} under the *GCM* and *CC-GCM*, a repeated measures ANOVA was conducted to examine the factors associated with relative bias. The estimating model was a within-subjects factor, and the multiple membership percentage and number of teachers encountered were between-subjects factors. ANOVA results revealed that relative bias depended on the estimating model and the percentage of multiple membership students in the data set (see Table 3).

Estimates of τ_{b00} were negatively biased under the *GCM* (RB (τ_{b00}) = -0.118) but positively biased under the *CC-GCM* (RB (τ_{b00}) = 0.073). Further, the magnitude of negative relative bias produced by the *GCM* decreased from -0.136 to -0.099 as the percentage of multiple membership students increased from 15% to 30%. By contrast, the magnitude of positive relative bias produced by the *CC-GCM* increased from 0.050 to 0.095 as the percentage of multiple membership students increased from 15% to 30%.

A second repeated measures ANOVA was conducted to examine the factors associated with relative bias for estimates of τ_{b10} under the *GCM* and *CC-GCM*. ANOVA results revealed that relative bias for estimates of τ_{b10} depended on the estimating model and two two-way interactions, as shown in Table 3. The magnitude of positive bias was significantly greater when the estimating model was *GCM* (RB (τ_{b10}) = 0.554) than it was when the estimating model was *CC-GCM* (RB (τ_{b10}) = 0.104). Further, the magnitude of positive relative bias produced by the *GCM* decreased from 0.590 to 0.519 as the multiple membership percentage increased from 15% to 30% of the students in the data set. By contrast, the magnitude of positive relative bias produced by the *CC-GCM* increased from 0.038 to 0.068 as the multiple membership percentage increased from 15% to 30%.

Similarly, the magnitude of positive relative bias produced by the *GCM* decreased from 0.564 to 0.544 as the number of teachers encountered by multiple membership students increased from two to three. The magnitude of positive relative bias produced by the *CC-GCM*, however, increased from 0.047 to 0.059 as the number of teachers encountered by multiple membership students increased from two to three.

Finally, a repeated measures ANOVA was conducted to examine the factors associated with relative bias for estimates of $\tau_{b00,10}$ under the *GCM* and *CC-GCM*. ANOVA results revealed that relative bias for estimates of $\tau_{b00,10}$ depended on the estimating model and an interaction between the estimating model and the percentage of multiple membership students in the data set. The magnitude of positive bias was significantly greater when the estimating model was *GCM* (RB ($\tau_{b00,10}$) = .596) than it was when the estimating model was *CC-GCM* (RB ($\tau_{b00,10}$) = .104). In addition, the magnitude of positive relative bias produced by the *GCM* decreased from 0.621 to 0.572 as the multiple membership percentage increased from 15% to 30%; whereas, the magnitude of positive relative bias produced by the *CC-GCM* increased from 0.082 to 0.126 across the same two conditions respectively.

To summarize, both the *GCM* and *CC-GCM* produced biased student random effects variance components estimates. The estimates produced by the *GCM*, however, were significantly more biased than those produced by the *CC-GCM*. Further, the multiple membership data structures tended to affect the two models differently. In general, as the percentage of multiple membership students in the data set and the number of teachers those students encountered increased, the magnitude of relative bias produced under the *GCM*, which ignores both cross-classified and multiple membership data structures, decreased. By contrast, as the percentage of multiple membership students in the data set and the number of teachers those students encountered increased, the magnitude of relative bias produced under the *CC-GCM*, which ignores only the multiple membership data structures, also increased.

Teacher random effects variance components

As described earlier in the paper, teacher effects in the data generated by the *CCMM-GCM* persist and accumulate across time. The effects for year-one teachers, year-two teachers, and year-three teachers are considered to be separate, additive teacher effects. Those teacher random effects estimates as produced by the *CC-GCM* and *CCMM-GCM* may be used to evaluate the effectiveness of each year's teachers (see e.g., Ponisciak & Bryk, (2005)). Therefore, to examine the teacher random effects variance estimates more carefully, we analyzed the estimates for the year-one, year-two, and year-three teachers separately under the *CCMM-GCM* and *CC-GCM*. The *GCM*, by contrast, models the students as purely nested within one classroom across time; therefore, the random effects are attributed to only the first-year teacher.

As expected, the correctly specified *CCMM-GCM* produced unbiased estimates of the year-one, year-two, and year-three teacher random effects variance components. By contrast, the estimates produced by the *GCM* and *CC-GCM* were significantly negatively biased. Given the identification of substantial bias for estimates of $\tau_{c00(1)}$ under the *GCM* and *CC-GCM*, a repeated measures ANOVA was conducted to examine the factors associated with relative bias. The ANOVA results revealed that the relative bias depended on the percentage of multiple membership students in the data and the number of teachers encountered by those students (see Table 4). In addition, there was a two-way interaction between the estimating model and the percentage of multiple membership students in the data.

The magnitude of negative relative bias for estimates of $\tau_{c00(1)}$ increased under both the *GCM* and *CC-GCM* as the percentage of multiple membership students in the data increased. When 15% of the students in the data set had multiple membership structures, the mean relative bias for estimates of $\tau_{c00(1)}$ was -0.154 and -0.145 under the *GCM* and *CC-GCM* respectively. On the other hand, when 30% of the students had multiple membership structures, the magnitude of the relative bias increased to -0.256 under the *GCM* and -0.296 under the CC-GCM. In other words, the magnitude of negative bias increased for both models as the percentage of multiple membership students in the data increased; however, the increase was larger under the *CC-GCM* than the *GCM*.

Table 4

Analysis of Variance Results for Mean Relative Bias of the Year-one Teacher Random Effect Variance Component, $\tau_{c00(1)}$, Year-two Teacher Random Effect Variance Component, $\tau_{c00(2)}$, and Year-three Teacher Random-effect Variance Component, $\tau_{c00(3)}$

	Parameter					
	$\tau_{c00(1)}$		$\tau_{c00(2)}$		$\tau_{c00(3)}$	
Source	*F-ratio*	$\eta^2_{partial}$	*F-ratio*	$\eta^2_{partial}$	*F-ratio*	$\eta^2_{partial}$
Between subjects						
mm% (P)	318.35*	**.138**	553.45*	**.217**	495.33*	**.199**
t (T)	52.00*	**.025**	69.99*	**.033**	145.70*	**.068**
P x T	9.72*	.005	13.92*	.006	1.02	< .001
Within subjects						
Model (M)	10.69*	.005	—	—	—	—
M x P	24.29*	**.012**	—	—	—	—
M x T	12.30*	.006	—	—	—	—

Note. — = not applicable. Practically significant $\eta^2_{partial}$ values ($\eta^2_{partial}$ > . 01) appear in bold face. **p* < .01.

Increasing the number of teachers encountered by multiple membership students had a similar effect on the mean relative bias produced by the two models. The negative bias produced by the *GCM* increased in magnitude from -0.188 to -0.222 as the number of teachers increased from two to three, and the negative bias produced by the *CC-GCM* increased in magnitude from -0.187 to -0.255.

A similar pattern was seen for estimates of the second- and third-year teacher random effects variance under the *CC-GCM*. ANOVA results revealed that the relative bias for estimates of $\tau_{c00(2)}$ and $\tau_{c00(3)}$ depended on the percentage of multiple membership students in the data and the number of teachers those students encountered (see Table 4). The magnitude of negative relative bias produced by the *CC-GCM* increased from -0.157 to -0.304 for estimates of $\tau_{c00(2)}$, and from -0.174 to -0.307 for estimates of $\tau_{c00(3)}$, as the percentage of multiple membership students in the data increased from 15% to 30%. Similarly, the magnitude of negative relative bias produced by the *CC-GCM* increased from -0.204 to -0.257 for estimates of $\tau_{c00(2)}$, and from -0.205 to -0.277 for estimates of $\tau_{c00(3)}$, as the number of teachers encountered by multiple membership students increased from two to three.

To summarize, failure to model multiple membership data structures within the data resulted in negatively biased estimates of the teacher random effect variance components. The magnitude of relative bias increased as the percentage of multiple membership students in the data and the number of teachers those students encountered increased.

Discussion

Failure to model cross-classified and multiple membership structures within the data did not seem to affect estimates of the fixed effects point estimates, although it should be noted that estimates of the intercept fixed effect were slightly outside the acceptable parameter bias range under the correctly specified *CCMM-GCM*.

Results indicate that the differing approaches to modeling cross-classified multiple membership growth data produce significantly different results in regard to random effect variance component estimates. This was particularly true in terms of the teacher and student random effects estimates with respect to achievement growth. The *CCMM-GCM*, which appropriately modeled the effects of multiple teachers on students' test scores and achievement growth, produced unbiased estimates of the student and teacher random effects variance. By contrast, the traditional *GCM* and the *CC-GCM* produced estimates that were biased.

The *GCM* and *CC-GCM* produced between-teacher variance estimates that were severely underestimated. This underestimation, in turn, resulted in the *GCM* and *CC-GCM* redistributing considerable teacher-level variance to the student level. This result is consistent with the findings of previous research suggesting that ignoring a crossed factor at the k^{th} level results in variability across level k units being redistributed to level k – 1 (see e.g., Luo & Kwok, 2009, 2012; Meyers & Beretvas, 2006).

Two factors contributing to the multiple membership data structure seemed related to this bias; a) the proportion of multiple membership students in the data, and b) the number of teachers those students encountered. Ignoring the relative contributions of multiple teachers to single test score outcomes appeared to cause the *GCM* and *CC-GCM* to underestimate the true variability of contributions between teachers. That underestimation became increasingly severe as the proportion of multiple membership students in the data and the number of teachers those students encountered increased. This result supports previous research finding that variance components of interest are underestimated when multiple membership data structures are ignored (Chung & Beretvas, 2011; Goldstein, 2010; Leckie, 2009).

The *CCMM-GCM* takes into account the relative contributions of the teacher effects to each student outcome. Given that the weights for the set of teachers associated with each observation at Level-1 sum to 1, it is well known that the variance resulting from a weighted sum of Level-2 residuals will always be smaller than the variance of the Level-2 residuals themselves (Goldstein, 2010; Leckie, 2009). The *GCM* and *CC-GCM*, however, do not reflect the reduced contribution of multiple teachers to the teacher random effects variance components, resulting in underestimation of the teacher random effects variance component. It is not surprising, therefore, that as the degree of multiple membership increased, so did the degree of underestimation bias identified for the teacher random effects variance components.

The results of this study have important implications for researchers and policy makers interested in estimating the effects that teachers have on student achievement growth. Results indicate that failure to properly account for the complex cross-classified multiple membership data structures that are likely to be seen in practice can result in considerably different estimates of teacher random effects variance components. It is important to note that the differences in the random effects estimates in this study would have led to decidedly different conclusions about the nature of teacher effects on student growth.

Measures used in teacher evaluation systems – including those that incorporate student growth data – are imperfect, which results in teachers being misclassified across different effectiveness categories. Considering that high stakes rewards and sanctions are increasingly associated with teacher effectiveness estimates, it is critically important for teacher evaluation systems to take every precaution to minimize such classification errors. The results of this study indicate that growth models which ignore the complexity of cross-classified multiple membership data structures would likely produce more classification errors than growth models which appropriately model that complexity. We therefore recommend that teacher evaluation systems use growth models that can accommodate the complexity of cross-classified multiple membership longitudinal data structures when using student growth data to estimate teacher effectiveness.

Limitations

This study was a preliminary investigation to assess how parameter estimates under the *CCMM-GCM* designed to handle cross classified multiple membership data structures differ from estimates under the use of a *GCM* that ignores cross classification and multiple membership, and a *CC-GCM* that ignores only multiple membership. Given that this was an initial attempt to quantify the potential biases, there were several limitations. Although the parameter values used in this simulation study were selected based on values seen in previous research, they represent a limited and somewhat simplified sample of values.

First, the current simulation study generated data to fit a simple, linear *CCMM-GCM* with only three measurement occasions. The data were balanced, meaning each student had the same number of measurements. Future research could examine parameter recovery when the data are unbalanced and/or non-linear growth trajectories are generated.

Second, and perhaps more importantly, the multiple membership data structures examined in this study represented a simplified version of the data structure seen in the real world data set. In particular, the number of teachers encountered by multiple membership students in this study was varied between two and three teachers. Furthermore, the multiple membership students had the same number of teachers at each measurement occasion. While this simplified data structure eased interpretation of the effects on relative bias, it was a poor representation of the structure seen in the real world, middle school data set. In that data set, multiple membership students encountered between two and eight teachers, and the number of teachers encountered fluctuated across measurement occasions. Students, for example, could have five teachers in year one, six teachers in year two, and eight teachers in year three.

Therefore, results should not be generalized before investigating more thoroughly how these factors affect *CCMM-GCM* estimation.

Although use of the *CCMM-GCM* under the examined conditions resulted in unbiased parameter estimates, future research should be conducted that assesses *CCMM-GCM* recovery of student and teacher random effects variance estimates under more realistic conditions. Given that residuals are increasingly being used to make important decisions about the value added by teachers and schools, more research is needed to assess methods that accurately portray and estimate the distributions of those residuals.

References

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three level hierarchical linear model. *American Journal of Education*, *97*, 65–108.

Chung, H., & Beretvas, S. N. (2012). The impact of ignoring multiple membership data structures in multilevel models. *British Journal of Mathematical and Statistical Psychology*, *65*(2), 185-200.

Dunlap, A. (2011). ESEA Flexibility: Department of Education Criteria and State Responses. *Mid-continent Research for Education and Learning (McREL)*.

Goldstein, H. (2010). *Multilevel statistical models* (4th ed.). New York: Hodder Arnold.

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*(4), 941-954.

Grady, M. W., & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivariate Behavioral Research*, *45*(3), 393-419.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, *26*(3), 329-367.

Hox, J. (2002). Multilevel analysis: Techniques and applications. Mahwah, NJ: Erlbaum.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*(4), 418-444.

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(3), 537-554.

Luo, W., & Kwok, O. M. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*, *44*(2), 182-212.

Luo, W., & Kwok, O. M. (2012). The Consequences of Ignoring Individuals' Mobility in Multilevel Growth Models A Monte Carlo Study. *Journal of Educational and Behavioral Statistics*, *37*(1), 31-56.

McCaffrey, D. F., Koretz, D., Lockwood, J. R., & Hamilton, L. S. (2004). *Evaluating value-added models for teacher accountability.* Accessed on January 9, 2014 from http://www.rand.org/pubs/monographs/MG158.

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research*, *41*(4), 473-497.

Murphy, D. L., & Pituch, K. A. (2009). The performance of multilevel growth curve models under an autoregressive moving average process. *The Journal of Experimental Education*, *77*(3), 255-284.

No Child Left Behind (NCLB). (2002). No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

Ponisciak, S. M., & Bryk, A. (2005). Value-added analysis of the Chicago public schools: An application of hierarchical models. *Value-added models in education: Theory and applications*, 40-79.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M. & Cameron, B. (2009) MLwiN Version 2.27. Centre for Multilevel Modelling, University of Bristol.

Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral statistics*, *19*(4), 337-350.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational and Behavioral Statistics*, *18*(4), 321-349.

Raudenbush, S.W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. Thousand Oaks, CA: Sage.

Version, S. A. S. 9.2 SAS Institute Inc. *SAS Campusl Drive, Cary, NC*, *27513*.