THE IMPACT OF MISSPECIFYING CROSS-CLASSIFIED RANDOM EFFECTS MODELS IN CROSS-SECTIONAL AND LONGITUDINAL MULTILEVEL DATA: A MONTE CARLO STUDY

A Dissertation

by

WEN LUO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2007

Major Subject: Educational Psychology

THE IMPACT OF MISSPECIFYING CROSS-CLASSIFIED RANDOM EFFECTS MODELS IN CROSS-SECTIONAL AND LONGITUDINAL MULTILEVEL DATA: A MONTE CARLO STUDY

A Dissertation

by

WEN LUO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by

Co-Chairs of Committee,	Victor Willson
	Oiman Kwok
Committee Members,	Bruce Thompson
	Michael Speed
Head of Department,	Michael Benz

August 2007

Major Subject: Educational Psychology

ABSTRACT

The Impact of Misspecifying Cross-Classified Random Effects Models in Cross-Sectional and Longitudinal Multilevel Data: A Monte Carlo Study. (August 2007)

Wen Luo, B.S., Shanghai Jiao Tong University;

M.A., Shanghai Jiao Tong University

Co-Chairs of Advisory Committee: Dr. Victor Willson
Dr. Oiman Kwok

Cross-classified random effects models (CCREMs) are used in the analyses of cross-sectional and longitudinal multilevel data that are not strictly hierarchical. Because of the complexity of this technique, many researchers simply ignore the cross-classified structures of their data and use hierarchical linear models. The study simulated cross-sectional and longitudinal multilevel data with cross-classified structures and examined the impact of misspecifying CCREMs on parameter and standard error estimates in these data.

The dissertation consists of two studies. Study One examines cross-sectional multilevel data and Study Two examines longitudinal multilevel data. In Study One, three-level cross-classified data were generated. Two random factors were crossed at either the top level or the intermediate level. It was found that ignoring a crossed random factor causes the variance of the remaining crossed factor and the adjacent levels to be overestimated. The fixed effects themselves are unbiased; however, the standard errors associated with the fixed effects are biased. When the ignored crossed factor is at the top

level, the standard error of the intercept is underestimated whereas the standard error of the regression coefficients associated with the covariate of the intermediate level and the remaining crossed factor are overestimated. When the ignored crossed factor is at the intermediate level, only the standard error of the regression coefficients associated with the covariate of the bottom level is overestimated.

In Study Two, longitudinal multilevel data were generated mirroring studies in which students are measured repeatedly and change schools over time. It was found that when the school level is modeled hierarchically above the student level rather than as a crossed factor, part of the variance at the school level is added to the student level, causing underestimation of the school-level variance and overestimation of the student-level variance and covariance. The standard errors of the intercept and the regression coefficients associated with the school-level predictors are underestimated, which may cause spurious significance for results.

The findings of the dissertation enhanced our understanding of the functioning of CCREMs in both cross-sectional and longitudinal multilevel data. The findings can help researchers to determine when CCREMs should be used and to interpret their results with caution when they misspecify CCREMs.

ACKNOWLEDGMENTS

I would like to thank my committee co-chairs, Dr. Willson and Dr. Kwok, and my committee members, Dr. Thompson and Dr. Speed, for their guidance and support throughout the course of the research.

Thanks to my friend Sanghan Lee for helping me with the syntax of the simulation. I also want to extend my gratitude to Dr. Hughes, who graciously gave me support and advice in my study and life. Thanks also to my colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my parents for their encouragement and to my fiancé for his patience and love.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

In educational and other social science research, multilevel data are quite commonly encountered. There are two types of multilevel data: hierarchical multilevel data and cross-classified multilevel data. In hierarchical multilevel data, the levels are strictly nested or hierarchical, in a sense that a lower-level unit or cluster belongs to one and only one higher-level cluster. For example, a student belongs to one and only one class and a class belongs to one and only one school. In cross-classified multilevel data, the data structure is not strictly hierarchical, but cross-classified. That is, the lower-level units are *cross*-classified by two or more higher-level factors, with each unit potentially belonging to any combination of the different factors. An example of cross-classified multilevel data occurs when students are cross-classified by the schools they attend and the neighborhoods they live in.

Cross-classified multilevel data can not only be found in cross-sectional studies, as in the example of students cross-classified by schools and neighborhoods, but also in longitudinal studies. For example, a number of students who are randomly drawn from a set of schools are tested repeatedly over time. Assuming students do not change schools over time, this dataset can be viewed as strictly hierarchical with three levels: occasions (or repeated measures) nested within students, and students nested within schools. However, if students change schools over time, the data are not strictly hierarchical any

---

This dissertation follows the style of *Structural Equation Modeling.*

more. In this case, occasions are cross-classified by students and schools.

In multilevel data, units belonging to the same cluster share the same cluster-specific influences. For example, students in the same class are taught by the same teachers and students in the same school use the same facilities. However, it is impossible to include all cluster-specific characteristics as covariates in an analysis. This is because we often have limited knowledge regarding relevant covariates, and our data may lack information on theses covariates. Therefore, there is cluster-level heterogeneity leading to dependence between responses of units in the same cluster after conditioning on covariates (Skrondal & Rabe-Hesketh, 2004).

In multilevel models, unobserved heterogeneity is modeled by including random effects in a multiple regression model. For strictly hierarchical multilevel data, hierarchical linear models (HLMs) are commonly used. For cross-classified multilevel data, cross-classified random effects models (CCREMs) were developed to accommodate grouping factors that are not nested (Goldstein, 1986, 1995; Raudenbush, 1993; Rasbash & Goldstein, 1994).

CCREMs provide much flexibility in modeling cross-classified multilevel data. In cross-sectional studies, CCREMs can model the effects of multiple contexts simultaneously (e.g., school, family, and neighborhood). In longitudinal studies, CCREMs can accommodate students' mobility and model school random effects more precisely. Despite the flexibility of CCREMs, researchers in substantive areas seldom use the technique because of its complexity or the lack of information about potential crossed factors in their data. A review of the published papers in educational research between 2004 and 2005 in the database of Education Resources Information Center (ERIC) showed that only *one* out of sixty studies with multilevel data used CCREMs. Many researchers simply ignored the cross-classified structures of their data and used hierarchical linear models.

Several simulation studies showed that ignoring a level of nesting in hierarchical linear models leads to biased variance component estimates and biased standard error estimates (Hutchison & Healy, 2001; Moerbeek, 2004; Opdenakker & Van Damme, 2000; Van Landeghem, De Fraine, & Van Damme, 2005). However, little research has been conducted to investigate the impact of ignoring cross-classified structures of multilevel data. The only introductory investigation was conducted by Meyers and his colleagues (Meyers, 2004; Meyers and Beretvas, 2006). It was found that if the cross-classified structure is ignored in a 2-level cross-classified data, the fixed effects estimates themselves are unaffected, however, the standard errors associated with the incorrectly modeled variables are underestimated; the bias of the standard errors gets worse as the variance attributable to the factor that was modeled incorrectly increases; and the level one variance and the variance of the remaining modeled crossed factor were overestimated.

Meyers and his colleagues only examined the functioning of 2-level CCREMs. The functioning of more general 3-level CCREMs is still unknown. In addition, little research has been conducted to investigate when it is necessary to use CCREMs in longitudinal multilevel data which have repeated measures cross-classified by individuals and clusters. The purpose of the dissertation is to systematically investigate the impact of misspecifying CCREMs as HLMs on parameter estimation for both cross-sectional and longitudinal cross-classified data.

The dissertation consists of five chapters. Chapter I introduces the background and states the purpose of the study. Chapter II reviews the specification and estimation of CCREMs. Chapter III presents the study that investigates the impact of misspecifying CCREMs in cross-sectional cross-classified data. Chapter IV presents the study that investigates the impact of misspecifying CCREMs in longitudinal cross-classified data. Chapter V summarizes the findings, discusses the implications of the findings, and provides directions for future research.

CHAPTER II

REVIEW OF LITERATURE

Cross-Classified Random Effects Models (CCREMs)

CCREMs for Cross-Sectional Data

Random Intercept CCREMs

There are two types of random effects, random intercepts and random coefficients. Random intercepts represent unobserved heterogeneity in the overall response, whereas random coefficients represent unobserved heterogeneity in the effects of explanatory variables on the response variable (Skrondal & Rabe-Hesketh, 2004). In this subsection, CCREMs with only random intercepts are introduced. In the following subsection, CCREMs with both random intercepts and random coefficients are introduced.

Consider the example of students being cross-classified by schools and neighborhoods. Let schools be indexed by j = 1, .., J, neighborhoods be indexed by k = 1, .., K, and students be indexed by i = 1, .., n_{jk}. The level-1 model without covariate is specified as

$$\text{Level-1: } y_{ijk} = \eta_{0jk} + \varepsilon_{ijk} \tag{2.1}$$

where η_{0jk} is the mean score specific to each cell (i.e., the cell mean score for students who attend the j^{th} school and live in the k^{th} neighborhood), and ε_{ijk} is the level-1 residual term.

In the level-2 model, the cell mean η_{0jk} is modeled as

$$\text{Level-2: } \eta_{0jk} = \gamma_{00} + \mu_{0j} + \nu_{0k} + \varpi_{0jk} \quad (2.2)$$

where γ_{00} is the grand mean, μ_{0j} is school random effect, ν_{0k} is the neighborhood random effect, and ϖ_{0jk} is the interaction effect.

The combined model is obtained by substituting level-2 model into level-1 model, yielding

$$y_{ijk} = \gamma_{00} + \mu_{0j} + \nu_{0k} + \varpi_{0jk} + \varepsilon_{ijk} . \quad (2.3)$$

This model resembles a conventional two-way analysis of variance (ANOVA) model except that the two main effects and the interaction effect are random instead of fixed.

In many applications, the within-cell sample sizes are not sufficiently large enough to distinguish the variance attributable to the interaction effect ϖ_{0jk} from that attributable to the residual ε_{ijk} (Raudenbush & Bryk, 2002). Therefore, the interaction effect is often collapsed to the level-1 residual, yielding

$$y_{ijk} = \gamma_{00} + \mu_{0j} + \nu_{0k} + \varepsilon_{ijk} . \quad (2.4)$$

Defining $\theta = Var(\varepsilon_{ijk})$, $\psi = Var(\mu_{0j})$, and $\tau = Var(\nu_{0k})$, it is typically assumed that $\varepsilon_{ijk} \sim i.i.d.N(0,\theta)$, $\mu_{0j} \sim N(0,\psi)$, and $\nu_{0k} \sim N(0,\tau)$. It is also assumed that the level-1 residual ε_{ijk}, the school random effect μ_{0j}, and the neighborhood random effect ν_{0k} are mutually independent. That is, $COV(\mu_{0j}, \varepsilon_{ijk}) = 0$, $COV(\nu_{0k}, \varepsilon_{ijk}) = 0$, and $COV(\mu_{0j}, \nu_{0k}) = 0$. The total variance of the outcome variable is composed of three variance components, the between school variance ψ, the between neighborhood variance τ, and the within-cell variance θ (i.e., $Var(y_{ijk}) = \psi + \tau + \theta$). Therefore, this model is also called the variance components model.

Students from the same school or neighborhood share the same environment, which results in potential dependency of the responses among students within the same school or neighborhood (i.e., students from the same school or neighborhood are more likely to have similar pattern of responses than students from different schools or neighborhoods). The dependency is measured by the correlation between two observations within a cluster, which is called intraclass correlation. Model (2.4) generates three kinds of intraclass correlation coefficients (ICCs): (1) the correlation between responses of two students who attend the same school and live in the same neighborhood: $Corr(y_{ijk}, y_{i'jk}) = \frac{\psi + \tau}{\psi + \tau + \theta}$; (2) the correlation between responses of two students who attend the same school but live in different neighborhoods: $Corr(y_{ijk}, y_{i'jk'}) = \frac{\psi}{\psi + \tau + \theta}$; (3) the correlation between responses of two students who live in the same neighborhood but attend different schools: $Corr(y_{ijk}, y_{i'j'k}) = \frac{\tau}{\psi + \tau + \theta}$.

For each source of the variability, we can attempt to explain it by including an explanatory variable associated with the source. A conditional random intercept model with a student-specific explanatory variable x_{ijk} (e.g., gender), a school-specific explanatory variable w_j (e.g., teacher student ratio), and a neighborhood-specific explanatory variable z_k (e.g., wealth) is specified as

$$y_{ijk} = \gamma_{00} + \gamma_{01}x_{ijk} + \gamma_{02}w_j + \gamma_{03}z_k + \mu_{0j} + \nu_{0k} + \varepsilon_{ijk} \qquad (2.5)$$

where γ_{00}, γ_{01}, γ_{02}, and γ_{03} are fixed coefficients, μ_{0j} is the residual school random effect, and ν_{0k} is the residual neighborhood random effect. The distributional assumptions of μ_{0j}, ν_{0k}, and ε_{ijk} are the same as those specified in model (2.4). The ICCs calculated based on the conditional random intercept model are called conditional ICCs.

Random Coefficient CCREMs

The random intercept model can be extended to a random coefficient model by allowing the regression coefficient of an explanatory variable to vary across clusters. For a random coefficient model with a single student-specific covariate x_{ijk}, the level-1 model is specified as

$$\text{Level-1: } y_{ijk} = \eta_{0jk} + \eta_{1jk}x_{ijk} + \varepsilon_{ijk} \qquad (2.6)$$

where η_{0jk} and η_{1jk} are the random intercept and regression coefficient, respectively. The variability of the intercept and the regression coefficient can be modeled as

$$\text{Level-2: } \eta_{0jk} = \gamma_{00} + \mu_{0j} + \nu_{0k}$$

$$\eta_{1jk} = \gamma_{10} + \mu_{1j} + \nu_{1k} \qquad (2.7)$$

where γ_{00} and γ_{10} are the overall intercept and regression coefficient, respectively. The school random effects (i.e., μ_{0j} and μ_{1j}) are specified as $\begin{bmatrix} \mu_{0j} \\ \mu_{1j} \end{bmatrix} \sim N\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} \psi_1 & \\ \psi_{12} & \psi_2 \end{bmatrix}$. The neighborhood random effects (i.e., ν_{0k} and ν_{1k}) are specified as $\begin{bmatrix} \nu_{0j} \\ \nu_{1j} \end{bmatrix} \sim N\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} \tau_1 & \\ \tau_{12} & \tau_2 \end{bmatrix}$.

It is typically assumed that μ_{0j} and μ_{1j} are uncorrelated with either ν_{0k} or ν_{1k}. That is, $COV(\mu_{0j}, \nu_{0k}) = 0$, $COV(\mu_{0j}, \nu_{1k}) = 0$, $COV(\mu_{1j}, \nu_{0k}) = 0$, and $COV(\mu_{1j}, \nu_{1k}) = 0$.

Substituting the level-2 model into the level-1 model yields the combined model

$$y_{ijk} = \gamma_{00} + \gamma_{10}x_{ijk} + \mu_{0j} + \nu_{0k} + \mu_{1j}x_{ijk} + \nu_{1k}x_{ijk} + \varepsilon_{ijk}. \quad (2.8)$$

Similar to the random intercept CCREMs, model (2.8) does not include a school by neighborhood interaction component, because generally the data matrix would be too sparse, that is, there would be too many empty cells, to estimate the interaction effect meaningfully. Some school-specific and neighborhood-specific predictors, such as teacher student ratio (w_j) and neighborhood wealth (z_k) could be added to model (2.7), giving

Level-2: $$\eta_{0jk} = \gamma_{00} + \gamma_{01}w_j + \gamma_{02}z_k + \mu_{0j} + \nu_{0k}$$

$$\eta_{1jk} = \gamma_{10} + \gamma_{11}w_j + \gamma_{12}z_k + \mu_{1j} + \nu_{1k}. \quad (2.9)$$

Substituting model (2.9) into model (2.6) yields the combined model

$$\begin{aligned} y_{ijk} &= \gamma_{00} + \gamma_{01}w_j + \gamma_{02}z_k + \gamma_{10}x_{ijk} + \gamma_{11}(w_j x_{ijk}) + \gamma_{12}(z_k x_{ijk}) \\ &\quad + \mu_{0j} + \nu_{0k} + \mu_{1j}x_{ijk} + \nu_{1k}x_{ijk} + \varepsilon_{ijk} \end{aligned}. \quad (2.10)$$

CCREMs could become very complex as more random effects are considered and more covariates are added. For example, in model (2.7) it is assumed that the effects of the school-specific predictor w_j on η_{0jk} and η_{1jk} do not vary across neighborhoods and the effects of the neighborhood-specific predictor z_k on η_{0jk} and η_{1jk} do not vary across schools. This model could be extended to allow randomly varying effects of school and neighborhood predictors. More complex CCREMs are not introduced because they are not used in the dissertation.

CCREMs for Longitudinal Data

Longitudinal data, often called panel data in social sciences, arise when units provide responses on multiple occasions. One important feature of longitudinal data is the clustering of responses within individuals. In longitudinal data analyses, a typical goal is to investigate the overall levels of the responses as well as changes in the responses over time.

Longitudinal data can be viewed as two-level data with repeated measures at level 1 nested within individuals at level 2. Individuals can be further nested in level-3 clusters, such as schools and therapists, resulting in longitudinal multilevel data. Such data can be analyzed using three-level hierarchical linear models in which the mean and covariance structures are typically modeled as a function of the time associated with the occasions (i.e., the time variable), the time-varying covariates, and the time-invariant covariates. One assumption of the hierarchical linear model is that individuals remain in the same cluster over time. When individuals move to different clusters, the hierarchical structure of the data is destroyed. Repeated measures become cross-classified by individuals and clusters.

Consider an example in which student's math achievement is measured annually for 4 years. Students are not strictly nested within schools because they move to different schools over time. Let students be indexed by $j = 1, .., J$, schools be indexed by $k = 1, ..,$

K, and occasions be indexed by t = 1, .., T. The level-1 model that assumes a straight-line growth trajectory in math scores for each student is specified as

$$\text{Level-1: } y_{tjk} = \pi_{0jk} + \pi_{1jk} x_{tjk} + \varepsilon_{tjk} \quad (2.11)$$

where x_{tij} is the time variable that measures the time elapsed between occasion t and the initial occasion. x_{tij} takes on value of 0, 1, 2, and 3 for year 1, 2, 3 and 4, respectively. π_{0jk} and π_{1jk} are the intercept and growth rate specific to the j^{th} student in the k^{th} school, and ε_{tjk} is the random residual that has the distribution $\varepsilon_{tjk} \sim i.i.d.N(0,\theta)$.

The level-2 model is specified as

$$\text{Level-2: } \pi_{0jk} = \gamma_{00} + \mu_{0j} + \nu_{0k}$$

$$\pi_{1jk} = \gamma_{10} + \mu_{1j} \quad (2.12)$$

where γ_{00} and γ_{10} are the overall intercept and growth rate, respectively. The μ_{0j} and μ_{1j} are the deviations of student j from the overall intercept and growth rate, respectively. They are distributed as $\begin{bmatrix} \mu_{0j} \\ \mu_{1j} \end{bmatrix} \sim N\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} \psi_1 & \\ \psi_{12} & \psi_2 \end{bmatrix}$. The ν_{0k} is the school random effect, which could be conceived as a deflection to a student's specific growth trajectory associated with studying in school k (Raudenbush & Bryk, 2002). The ν_{0k} is normally distributed with mean 0 and variance τ.

The combined model is obtained by substituting model (2.12) into model (2.11), yielding

$$y_{tjk} = \gamma_{00} + \mu_{0j} + \nu_{0k} + (\gamma_{10} + \mu_{1j})x_{tjk} + \varepsilon_{tjk}\,. \qquad (2.13)$$

Consider a hypothetical case in which student *j* attends school 1 at occasion 1, school 2 at occasion 2, and school 3 at occasion 3. The predicted values for that student, given that student's growth parameters and the school effects, would be $\hat{y}_{1j1} = \gamma_{00} + \mu_{0j} + \nu_{01}$ at occasion 1; $\hat{y}_{2j2} = \gamma_{00} + \mu_{0j} + \nu_{02} + \gamma_{10} + \mu_{1j}$ at occasion 2; and $\hat{y}_{3j3} = \gamma_{00} + \mu_{0j} + \nu_{03} + 2(\gamma_{10} + \mu_{1j})$ at occasion 3.

The growth trajectory of the student is displayed in Figure 1. The solid line represents the predicted growth trajectory. The first dashed line represents the student's trajectory if the student did not change from school 1 to school 2 and the second dashed line represents the trajectory if he/she did not change from school 2 to school 3. The gain from occasion 1 to occasion 2 would be $\gamma_{10} + \mu_{1j} + \nu_{02} - \nu_{01}$ and the gain from occasion 2 to occasion 3 would be $\gamma_{10} + \mu_{1j} + \nu_{03} - \nu_{02}$.

In model (2.13), it is assumed that the effect of a school would disappear when students move out of the school. In other words, previous schools would not affect student's outcome in current schools. There are other CCREMs that specify cumulative school effects (McCaffrey, Lockwood, Koretz, & Hamilton, 2004; Raudenbush & Bryk, 2002). These models, known as value-added models (VAMs), are often used in longitudinal student achievement data linked to teachers and schools to make inferences about teacher and school effectiveness (Lockwood, Doran, & McCaffrey, 2003). The details of value-added cross-classified models are not be elaborated upon because the dissertation only uses CCREMs without cumulative cluster effects.

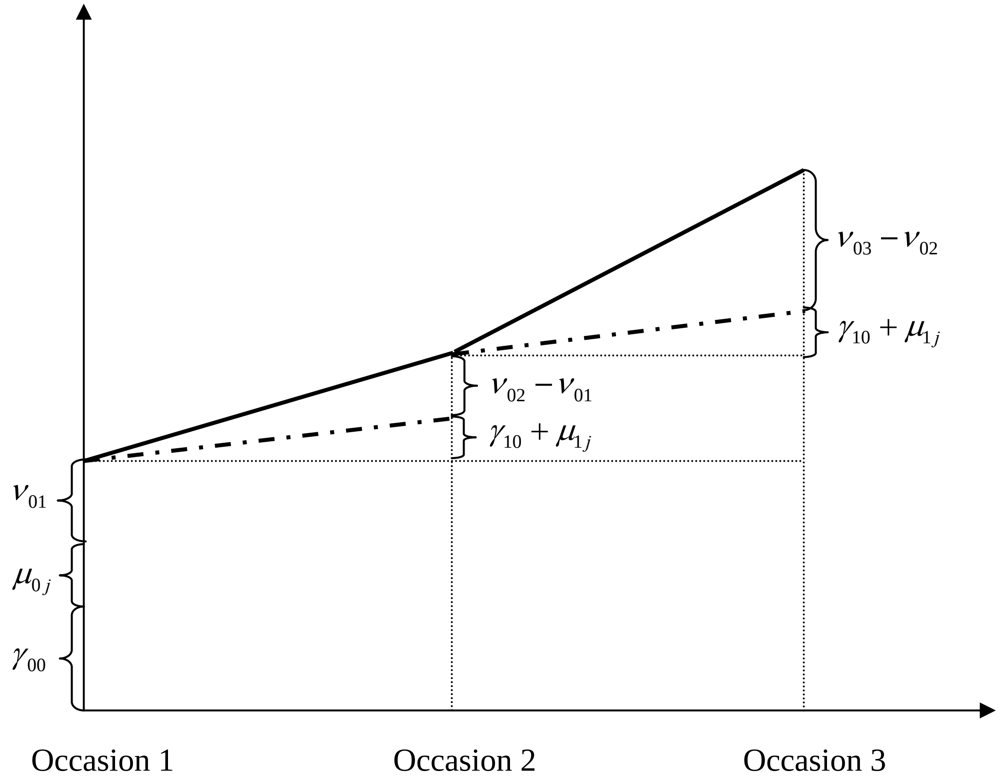

FIGURE 1 The growth trajectory of student *j* attending different schools over time

Estimation of Cross-Classified Random Effects Models

A variety of approaches have been developed to estimate cross-classified random effects models. This section introduces three commonly used likelihood-based approaches.

Estimation of CCREMs as Mixed Linear Models

CCREMs can be viewed as a special case of mixed linear models. The general mixed linear model is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad (2.14)$$

where $\mathbf{y}$ denotes the vector of observed y_i's, $\mathbf{X}$ is the known matrix of covariates, $\boldsymbol{\beta}$ is the vector of unknown fixed-effects parameters, $\mathbf{Z}$ is the known design matrix, $\mathbf{u}$ is the vector unknown random-effects parameters, and $\mathbf{e}$ is the unobserved vector of random errors. It is assumed that $\mathbf{u}$ and $\mathbf{e}$ are normally distributed with $E\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$ and $Var\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & 0\\ 0 & \mathbf{R}\end{bmatrix}$. The variance of $\mathbf{y}$ is therefore $\mathbf{V}=\mathbf{ZGZ}'+\mathbf{R}$.

Suppose there are 12 students cross-classified by 2 schools and 3 neighborhoods as is shown in Table 1. A mixed model can be written out as

$$\begin{bmatrix} y_1\\ y_2\\ y_3\\ y_4\\ y_5\\ y_6\\ y_7\\ y_8\\ y_9\\ y_{10}\\ y_{11}\\ y_{12} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p}\\ x_{21} & x_{22} & \cdots & x_{2p}\\ x_{31} & x_{32} & \cdots & x_{3p}\\ x_{41} & x_{42} & \cdots & x_{4p}\\ x_{51} & x_{52} & \cdots & x_{5p}\\ x_{61} & x_{62} & \cdots & x_{6p}\\ x_{71} & x_{72} & \cdots & x_{7p}\\ x_{81} & x_{82} & \cdots & x_{8p}\\ x_{91} & x_{92} & \cdots & x_{9p}\\ x_{101} & x_{102} & \cdots & x_{10p}\\ x_{111} & x_{112} & \cdots & x_{11p}\\ x_{121} & x_{122} & \cdots & x_{12p} \end{bmatrix} \times \begin{bmatrix} \beta_1\\ \beta_2\\ \beta_3\\ \vdots\\ \beta_p \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 & 0 & 0\\ 1 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0 & 1\\ 1 & 0 & 0 & 0 & 1\\ 0 & 1 & 1 & 0 & 0\\ 0 & 1 & 1 & 0 & 0\\ 0 & 1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1 & 0\\ 0 & 1 & 0 & 0 & 1\\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} u_1\\ u_2\\ v_1\\ v_2\\ v_3 \end{bmatrix} + \begin{bmatrix} e_1\\ e_2\\ e_3\\ e_4\\ e_5\\ e_6\\ e_7\\ e_8\\ e_9\\ e_{10}\\ e_{11}\\ e_{12} \end{bmatrix} \tag{2.15}$$

where u_1 and u_2 are the random effects associated with School 1 and 2, respectively; v_1, v_2, and v_3 are the random effects associated with Neighborhood 1, 2, and 3, respectively. Let ψ denote the variance of school random effects, τ the variance of neighborhood random effects, and θ the residual variance. Then **G** and **R** have the following forms:

$$\mathbf{G} = \begin{bmatrix} \psi & 0 & 0 & 0 & 0 \\ 0 & \psi & 0 & 0 & 0 \\ 0 & 0 & \tau & 0 & 0 \\ 0 & 0 & 0 & \tau & 0 \\ 0 & 0 & 0 & 0 & \tau \end{bmatrix} \text{ and } \mathbf{R} = \begin{bmatrix} \theta & 0 & 0 & \cdots & 0 \\ 0 & \theta & 0 & \cdots & 0 \\ 0 & 0 & \theta & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \theta \end{bmatrix}.$$

TABLE 1

Twelve Students Cross-Classified by Two Schools and Three Neighborhoods

	School 1	School 2
Neighborhood 1	Student 1, 2	Student 7, 8
Neighborhood 2	Student 3, 4	Student 9,10
Neighborhood 3	Student 5, 6	Student 11, 12

In SAS PROC MIXED (SAS Institute Inc., 2004), the default estimation method for parameters in **G** and **R** is restricted maximum likelihood (REML), which has the log likelihood function:

$$l_R(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n-p}{2}\log 2\pi \qquad (2.16)$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and p is the rank of **X** (Littell, Milliken, Stroup, & Wolfinger, 1996). The fixed-effect parameters $\boldsymbol{\beta}$ are estimated by generalized least-squares (GLS), minimizing $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

There are different algorithms to minimize $-2l_R(\mathbf{G},\mathbf{R})$. PROC MIXED implements a ridge-stabilized Newton-Raphson algorithm and sweep-based algorithm

(Wolfinger, Tobias, & Sall, 1994). Bates (2004, 2005) presented the sparse matrix algorithm that is implemented in the **lmer** function in the R package **lme4**.

The Raudenbush Approach

Raudenbush (1993) proposed a way to estimate cross-classified random effects models by combining the concepts of exchangeability between and within regressions. Exchangeability between regressions means that the estimation of a regression equation is replicated across a number of similar units, such as neighborhoods, and there is no prior basis to predict how the regression coefficients of neighborhood *r* differ from those of neighborhood r'. Exchangeability within regressions means that in a single regression equation that includes many predictors that are measured on the same scale, there is no prior belief regarding the relative magnitude of the regression coefficients of those predictors so that a prior probability distribution can be assigned to those unknown coefficients.

Consider the example of students being cross-classified by *J* schools and *K* neighborhoods. Regressions will be exchangeable between the *J* schools whereas indicator variables for the *K* neighborhoods will have exchangeable coefficients within each of the *J* regressions. For the j^{th} school, the cross-classified model is written as

$$\mathrm{Y}_j = \mathrm{X}_j^{(1)}\beta^{(1)} + \mathrm{Z}_j^{(1)}\zeta_j^{(1)} + \Lambda_j^{(2)}\mathrm{X}^{(2)}\beta^{(2)} + \Lambda_j^{(2)}\zeta^{(2)} + \mathrm{E}_j \tag{2.17}$$

where matrices with the superscript (1) are defined on students and schools whereas matrices with the superscript (2) are defined on neighborhoods. $\mathrm{X}_j^{(1)}$ and $\mathrm{X}^{(2)}$ are the design matrices for the fixed-effects vectors $\beta^{(1)}$ and $\beta^{(2)}$, respectively. $\mathrm{Z}_j^{(1)}$ is the design matrix for the random-effects vector $\zeta_j^{(1)}$. The covariance matrix of $\zeta_j^{(1)}$ is denoted as Ψ. The Λ_j is a n_j by *K* matrix of indicators with columns *k* = 1, .., *K* corresponding to neighborhoods. The element $\Lambda_j(i,k) = 1$ if student *i* lives in neighborhood *k* and $\Lambda_j(i,k) = 0$ otherwise. The $\zeta^{(2)}$ is a *K* by 1 vector of random effects defined on neighborhoods and its covariance matrix is a diagonal matrix with a common variance δ of the neighborhood-level random effects on the diagonal. E_j is a vector of residual that has a normal distribution with mean 0 and variance θ.

Using the example shown in Table 1, the matrix presentation of a CCREM could be written out as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1p}^{(1)} \\ x_{21}^{(1)} & x_{22}^{(1)} & \cdots & x_{2p}^{(1)} \\ x_{31}^{(1)} & x_{32}^{(1)} & \cdots & x_{3p}^{(1)} \\ x_{41}^{(1)} & x_{42}^{(1)} & \cdots & x_{4p}^{(1)} \\ x_{51}^{(1)} & x_{52}^{(1)} & \cdots & x_{5p}^{(1)} \\ x_{61}^{(1)} & x_{62}^{(1)} & \cdots & x_{6p}^{(1)} \\ x_{71}^{(1)} & x_{72}^{(1)} & \cdots & x_{7p}^{(1)} \\ x_{81}^{(1)} & x_{82}^{(1)} & \cdots & x_{8p}^{(1)} \\ x_{91}^{(1)} & x_{92}^{(1)} & \cdots & x_{9p}^{(1)} \\ x_{101}^{(1)} & x_{102}^{(1)} & \cdots & x_{10p}^{(1)} \\ x_{111}^{(1)} & x_{112}^{(1)} & \cdots & x_{11p}^{(1)} \\ x_{121}^{(1)} & x_{122}^{(1)} & \cdots & x_{12p}^{(1)} \end{bmatrix} \times \begin{bmatrix} \beta_1^{(1)} \\ \beta_2^{(1)} \\ \vdots \\ \beta_p^{(1)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} \zeta_1^{(1)} \\ \zeta_2^{(1)} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1q}^{(2)} \\ x_{21}^{(2)} & x_{22}^{(2)} & \cdots & x_{2q}^{(2)} \\ x_{31}^{(2)} & x_{32}^{(2)} & \cdots & x_{3q}^{(2)} \end{bmatrix} \times \begin{bmatrix} \beta_1^{(2)} \\ \beta_2^{(2)} \\ \vdots \\ \beta_q^{(2)} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \zeta_1^{(2)} \\ \zeta_2^{(2)} \\ \zeta_3^{(2)} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \end{bmatrix} . \tag{2.18}$$

Equation (2.17) can be further simplified by consolidating the fixed effects part of the model, yielding

$$\mathrm{Y}_j = \mathrm{X}_j \beta + \mathrm{Z}_j^{(1)} \zeta_j^{(1)} + \Lambda_j^{(2)} \zeta^{(2)} + \mathrm{E}_j \tag{2.19}$$

where $X_j = [X_j^{(1)} \mid \Lambda_j^{(2)} X^{(2)}]$ and $\beta = \begin{bmatrix} \beta^{(1)} \\ \beta^{(2)} \end{bmatrix}$. Equation (2.19) can be viewed as a special case of the general mixed linear model

$$Y_j = X_j \beta + Z_j \zeta_j + E_j \tag{2.20}$$

where $Z_j = [Z_j^{(1)} \mid \Lambda_j^{(2)}]$ and $\zeta_j = \begin{bmatrix} \zeta_j^{(1)} \\ \zeta^{(2)} \end{bmatrix}$. The covariance matrix of ζ_j becomes $\begin{bmatrix} \Psi & 0 \\ 0 & \delta \times I \end{bmatrix}$.

Being transformed to a general mixed linear model, a CCREM is estimated using REML method with the Expectation-Maximization (EM) algorithm (Dempster, Rubin, & Tsutakawa, 1981).

The Raudenbush approach is implemented in HLM 6 (Raudenbush, Bryk, Cheong, & Congdon, 2004). This approach is most appropriate for data with two-way cross-classification where one random factor contains a large number of units and the other a comparatively small number of units. Usually, the factor with a large number of units is treated as the primary hierarchical one whereas the factor with fewer units is added into the hierarchy by assigning each unit an indicator variable with a random coefficient.

The Rasbash and Goldstein Approach

Rasbash and Goldstein (1994) described a likelihood-based approach that involves transforming a CCREM into a constrained hierarchical model that is then estimated using the standard iterative generalized least squares (IGLS) algorithm.

Consider the example of students cross-classified by *J* schools and *K* neighborhoods. First, a 'virtual' level (i.e., level-3) within which both schools and neighborhoods are nested (for instance towns) is introduced. Note that the virtual level does not need to be a natural level. It could be defined as a single unit encompassing the whole dataset if it is not possible to find a natural third level. Then at the student level, dummy variables d_{ik} (k = 1, .., K) are specified to indicate neighborhoods. The d_{ik} equals 1 if student *i* lives in neighborhood *k* and zero otherwise. Using these dummy variables, we can write model (2.5) equivalently as

$$y_{ijk} = \mathbf{X\beta} + \mathbf{Zu} + \sum_{k=1}^{K} v_{0k} d_{ik} + \varepsilon_{ijk} \tag{2.21}$$

where v_{0k} is the random regression coefficient associated with d_{ik} that varies at level-3. v_{0k} is assumed to be normally distributed with mean zero and variance τ. The covariance of v_{0k} is constrained to be zero because neighborhoods are assumed to be mutually independent. The covariance between v_{0k} and the school random effects (i.e., $\mathbf{u}$) are also constrained to be zero.

By introducing an explanatory variable x_k with random coefficient, model (2.21) can be extended to

$$y_{ijk} = \mathbf{X\beta} + \mathbf{Zu} + \sum_{k=1}^{K} (v_{0k} d_{ik} + v_{1k} d_{ik} x_k) + \varepsilon_{ijk} . \tag{2.22}$$

The joint distribution of ν_{0k} and ν_{1k} is assumed to be $\begin{bmatrix} \nu_{0k} \\ \nu_{1k} \end{bmatrix} \sim N\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} \tau_1 & \\ \tau_{12} & \tau_2 \end{bmatrix}$. Being transformed to a hierarchical model, the CCREM is then fitted using iterative generalized least square (IGLS) algorithm (Goldstein, 1986).

The Rasbash and Goldstein approach is similar to the Raudenbush approach in that both methods treat one of the crossed random factors as a primary hierarchical one and incorporate the other factor by assigning indicator variables with random coefficients to the units. Unlike the Raudenbush approach, the Rasbash and Goldstein approach is also appropriate for models with more than 2 crossed random factors and with more levels. The Rasbash and Goldstein approach is implemented in MLwiN 2.0 (Rasbash, Steele, Browne, & Prosser, 2004).

CHAPTER III

STUDY ONE: IGNORING CROSS-CLASSIFIED STRUCTURES IN CROSS-SECTIONAL DATA

Meyers (2004) conducted a simulation study investigating the impact of ignoring a crossed factor in cross-classified data. They generated 2-level cross-classified data in which level-1 units were cross-classified by two level-2 factors (i.e., students cross-classified by middle schools and high schools). Student's gender was included in the model as a level-1 predictor, the size of middle schools as a predictor associated with the random factor of middle schools, and the percentage of the students in the free or reduced lunch program at high schools as a predictor associated with the random factor of high schools.

The generated data were analyzed using two models: the correct model (i.e., cross-classified model) and the misspecified model (i.e., hierarchical linear model). In the misspecified model, the random factor of middle schools was ignored and the corresponding covariate (i.e., the size of middle schools) was modeled as a *level-1* (i.e., student-level) predictor.

They found that the estimated intercept and regression coefficients were unaffected; however, the standard error of the regression coefficient associated with the inappropriately modeled covariate was underestimated (i.e., smaller than the true standard error value). Moreover, the level-1 residual variance (i.e., variance between students) and the variance of the remaining modeled crossed factor (i.e., variance between high schools) were overestimated.

The design factors that they manipulated included: the correlation between the two crossed factors (i.e., middle schools and high schools), the number of middle schools feeding into each high school[1], the number of middle schools and high schools, the average number of students sampled from each middle school, and the middle school and high school intraclass correlation. The intraclass correlation of the crossed factors and the correlation between the crossed factors were found to have large effects on the magnitude of the bias. The average number of students per middle school and the number of feeder middle schools were found to have small effects on the standard error bias.

Meyers only examined the functioning of CCREMs for two-level cross-classified data. Although 2-level cross-classified data are very common in educational research, the findings of Meyers' study cannot be generalized to cross-classified data with more than two levels. It is desirable to examine more general 3-level cross-classified data to understand the performance of CCREMs better. Theoretically crossed random factors can occur at any higher level in a multilevel data set (Hox, 2002). Crossed random factors can occur at the top level. For example, in meta-analysis, effect sizes are nested within studies and studies are cross-classified by authors and samples[2]. The data structure is presented in Figure 2 using the classification diagram (Rasbash & Browne, 2006). Crossed random factors can also occur at the intermediate level. For example,

---

[1] This controls the degree of cross-classification or the number of empty cells in the data set. The fewer the number of middle school feeders, the less crossed the two factors are and the more empty cells exist in the data set.

[2] Studies are cross-classified by authors and samples because one author could conduct different studies using different samples. Similarly, there could be studies conducted by different authors with exactly the same sample (e.g., studies using archival data).

students are cross-classified by classes and study groups, and both classes and study groups are nested within schools (see Figure 3).

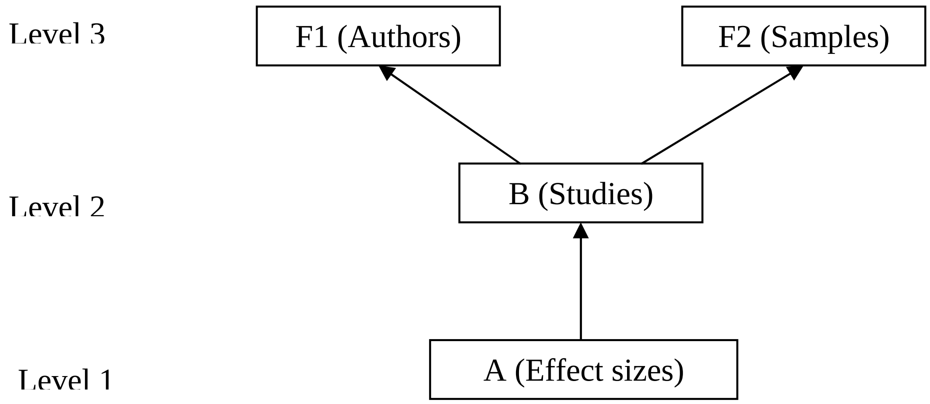

FIGURE 2 Crossed random factors at the highest level

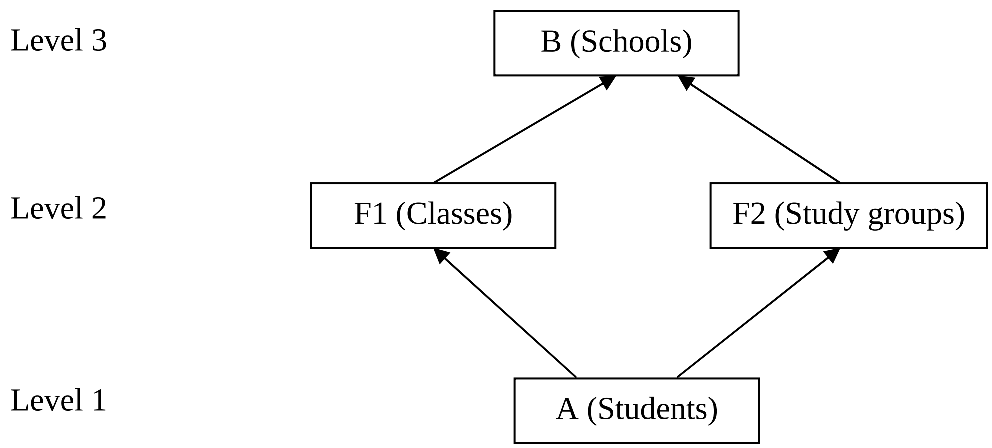

FIGURE 3 Crossed random factors at the intermediate level

Moerbeek (2004) investigated the consequences of ignoring a level of nesting in a 3-level hierarchical linear model and found that whether a parameter estimate is biased depends on which level is ignored (i.e. the top level or the intermediate level). Therefore, it is important to investigate the impact of ignoring a crossed factor in both situations (i.e., crossed factors at the top level vs. at the intermediate level) using 3-level cross-classified data.

In Meyers' study, the magnitudes of the intraclass correlation coefficients were constrained to be the same for the two crossed factors and the cluster sizes of the two crossed factors were also set to be equal. It is of interest to add conditions of uneven intraclass correlation coefficients and uneven cluster sizes for the two crossed factors.

To provide a better understanding of the impact of misspecifying CCREMs in cross-classified data, the present study expanded Meyers' study in two ways: (1) generating 3-level cross-classified data in which misspecifications could occur at either the intermediate or the top level, and (2) adding conditions of unequal intraclass correlation coefficients and unequal cluster sizes for the two crossed factors. Two simulations were conducted. In Simulation One, two random factors were crossed at the top level whereas in Simulation Two, they were crossed at the intermediate level. Results were presented to show how the estimated variance of the random effects, the estimated fixed effects and the corresponding estimated standard errors were influenced under a variety of manipulated conditions.

Simulation One

Methods

Data Generation

Data were generated based on a three-level random intercept cross-classified model with two random factors crossed at the top level (i.e., Level-3 in Figure 1). The model is specified as:

$$y = \gamma_0 + \gamma_1 x + \mu + \nu + \omega + \varepsilon \qquad (3.1)$$

where x is a covariate associated with either level-1 unit A, level-2 random factor B, or level-3 random factor F1. Covariates associated with the other level-3 random factor F2 (i.e., the ignored factor in the misspecified model) were not included in the model for two reasons. First, it is common for researchers to ignore a crossed random factor and the covariates associated with the factor as well. When the group membership data are absent, the covariates associated with the grouping factor are very likely to be unavailable (Moerbeek, 2004). Second, in practice there are datasets that contain covariates associated with a certain grouping factor without the information of the group membership (Van Landeghem, De Fraine, & Van Damme, 2005). However, because Meyers and Beretvas (2006) have investigated this situation (i.e., ignoring a crossed factor but keeping the associated covariate by treating it as a lower level predictor), the present study only focuses on the situation in which both the crossed random factor and its covariates are ignored.

In model (3.1), γ_0 and γ_1 are the fixed intercept and regression coefficient, respectively. μ is the random effect of the level-2 factor B, which is specified as

$\mu \sim N(0,\psi)$; v is the random effect associated with the level-3 factor F1, which is specified as $v \sim N(0,\tau)$; and ω is the random effect associated with the level-3 factor F2, which is specified as $\omega \sim N(0,\varsigma)$. ε is the level-1 residual, which is specified as $\varepsilon \sim N(0,\theta)$.

In this simulation, γ_0 and γ_1 were set to be .10 and .50, respectively, for all conditions. ψ (i.e., variance of the level-2 factor B) and θ (i.e., level-1 residual variance) were set to be .2 and .6, respectively. The magnitudes of τ (i.e., variance of the level-3 factor F1) and ς (i.e., variance of the level-3 factor F2) were manipulated under different conditions.

The total number of the level-1 units A was 4000 (i.e., N = 4000). For the balanced design, the 4000 level-1 units A were nested within 500 clusters of the level-2 factor B, and the cluster size of factor B was 8 (i.e., $n_B = 8$ observations per cluster) . The 500 clusters of the level-2 factor B were cross-classified by the two level-3 factors F1 and F2. The cluster size of F1 and F2 (i.e., n_{F1} and n_{F2}) were manipulated under different conditions.

Design Factors

The design factors that were manipulated included: the conditional intraclass correlation coefficients of F1 and F2 (i.e., ICC_{F1} and ICC_{F2}) , the cluster size of F1 and F2 (i.e., n_{F1} and n_{F2}) , and the degree of cross-classification.

Intraclass correlation coefficient. Based on model (3.1), the conditional ICC of F1 is computed by $ICC_{F1} = \frac{\tau}{\theta + \psi + \tau + \varsigma}$ and the conditional ICC of F2 is computed by $ICC_{F2} = \frac{\varsigma}{\theta + \psi + \tau + \varsigma}$. There were three levels in this design factor. First, F1 had larger ICC than F2: $ICC_{F1} = .15$ and $ICC_{F2} = .05$. Second, F1 and F2 had equal ICC: $ICC_{F1} = ICC_{F2} = .10$. Third, F1 had smaller ICC than F2: $ICC_{F1} = .05$ and $ICC_{F2} = .15$.

The intraclass correlation coefficients are directly related to the magnitude of the conditional variances of the two crossed factors. For instance, in the first condition where $ICC_{F1} = .15$ and $ICC_{F2} = .05$, given that the total conditional variance of y was 1.0 (i.e., $V(y \mid x) = 1.0$), ICC_{F1} = .15 indicated that the conditional variance of F1 was .15.

Cluster size. For the balanced design, the cluster size of a level-3 crossed factor was defined as the number of level-2 units in each cluster of the factor. There were three levels of cluster size. First, F1 had greater cluster size than F2: $n_{F1} = 25$ and $n_{F2} = 10$. Given that the total number of the level-2 units was 500, the total number of clusters in F1 was 20 (i.e., 500/25 = 20, with 25 level-2 units per cluster) and the total number of clusters in F2 was 50 (i.e., 500/10 = 50, with 10 level-2 units per cluster). Second, F1 and F2 had equal cluster size: $n_{F1} = n_{F2} = 20$; that is, the number of clusters in F1 = the number of clusters in F2 = 25. Third, F1 had smaller cluster size than F2: $n_{F1} = 10$ and

$n_{F2} = 25$; that is, the number of clusters in F1 = 50 and the number of clusters in F2 = 20.

Degree of cross-classification. The degree of cross-classification controls the distribution of units among the crossed random factors. Table 2 presents the situation of a full cross-classification (i.e., %Cross = 100%) of two random factors F1 and F2. Suppose there are 4 clusters (e.g., 4 schools) in F1 and 6 clusters (e.g., 6 neighborhoods) in F2. ρ_{ij} (i = 1, .., 4, j = 1, .., 6) is the probability of an observation belonging to cluster-*i* of F1 and cluster-*j* of F2. In full cross-classification, ρ_{ij} is greater than 0 for any combination of *i* and *j*. In other words, observations in a specific cluster of F1 can belong to any clusters of F2 and observations in a specific cluster of F2 can also belong to any clusters of F1.

Table 3 presents the fully nested situation in which F2 is nested within F1 (i.e., %Cross = 0). Observations in cluster 1 and 2 of F2 can only belong to cluster 1 of F1 and observations in cluster 3 and 4 of F2 can only belong to cluster 2 of F1. Table 4 presents the situation of partial cross-classification of F1 and F2. Observations in cluster 1, 2 and 3 of F2 can belong to either cluster 1 or 2 of F1, but they cannot belong to cluster 3 or 4 of F1. On the contrary, observations in cluster 4, 5 and 6 of F2 can belong to either cluster 3 or 4 of F1, but they cannot belong to cluster 1 or 2 of F1. This situation is called as 50% cross-classification (i.e., %Cross = 50%), because only half of the clusters in F1 and half of the clusters in F2 are fully crossed.

In reality, partial cross-classification is very common. For example, students living in certain neighborhoods only go to certain schools and students attending certain schools only live in certain neighborhoods. In this study, there were two levels in the degree of cross-classification: (1) the two level-3 factors, F1 and F2, were fully crossed (%Cross = 100%), and (2) the two level-3 factors were partially crossed (%Cross = 50%).

Combining the three design factors, the simulation involved a total of 18 design conditions (i.e., 3 ICCs × 3 Cluster sizes × 2 Degrees of cross-classification). For each condition, 500 datasets were generated. Each dataset was then analyzed using two models: (1) the correctly specified model (i.e., the cross-classified model), in which both crossed factors were included; (2) the misspecified model (i.e., the hierarchical linear model), in which one of the crossed factors (i.e., F2) was ignored.

The data were generated using SAS 9.1 and the models were then estimated using PROC MIXED with restricted maximum likelihood estimation method (SAS Institute Inc., 2004).

Analysis

The fixed and random effect parameter estimates from the correctly specified model and the misspecified model were summarized across the 500 replications for each of the 18 conditions. The relative bias of parameter estimates was calculated using the following formula:

$$B(\hat{\theta}) = \frac{\bar{\hat{\theta}} - \theta}{\theta} \tag{3.4}$$

where $\bar{\hat{\theta}}$ is the mean of a parameter estimate across the 500 replications and θ is the true parameter value. A negative relative bias indicates an underestimation of the

TABLE 2

Full Cross-Classification of F1 and F2 (%Cross = 100%)

		F2					
		1	2	3	4	5	6
F1	1	ρ_{11}	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{15}	ρ_{16}
	2	ρ_{21}	ρ_{22}	ρ_{23}	ρ_{24}	ρ_{25}	ρ_{26}
	3	ρ_{31}	ρ_{32}	ρ_{33}	ρ_{34}	ρ_{35}	ρ_{36}
	4	ρ_{41}	ρ_{42}	ρ_{43}	ρ_{44}	ρ_{45}	ρ_{46}

TABLE 3

F2 Nested Within F1 (%Cross = 0%)

		F2					
		1	2	3	4	5	6
F1	1	ρ_{11}	ρ_{12}	0	0	0	0
	2	0	0	ρ_{23}	ρ_{24}	0	0
	3	0	0	0	0	ρ_{35}	0
	4	0	0	0	0	0	ρ_{46}

TABLE 4

Partial Cross-Classification of F1 and F2 (%Cross = 50%)

		F2					
		1	2	3	4	5	6
F1	1	ρ_{11}	ρ_{12}	ρ_{13}	0	0	0
	2	ρ_{21}	ρ_{22}	ρ_{23}	0	0	0
	3	0	0	0	ρ_{34}	ρ_{35}	ρ_{36}
	4	0	0	0	ρ_{44}	ρ_{45}	ρ_{46}

parameter (i.e., the estimated value is smaller than the true parameter value), whereas a positive relative bias indicates an overestimation of the parameter (i.e., the estimated value is larger than the true parameter value). Using the cutoff value recommended by Hoogland and Boomsma (1998), relative bias that has an absolute value less than .05 was considered acceptable.

The relative bias of estimated standard errors was computed using the following formula:

$$B(\hat{S}_{\hat{\theta}}) = \frac{\bar{\hat{s}}_{\hat{\theta}\_MIS} - \bar{\hat{s}}_{\hat{\theta}\_CCM}}{\bar{\hat{s}}_{\hat{\theta}\_CCM}} \qquad (3.5)$$

where $\bar{\hat{S}}_{\hat{\theta}\_MIS}$ is the mean of the estimated standard error across the 500 replications based on the misspecified model (i.e., the hierarchical linear model) and $\bar{\hat{S}}_{\hat{\theta}\_CCM}$ is the mean estimated standard error based on the correct model (i.e., the cross-classified model). Hoogland and Boomsma (1998) recommended that the relative bias of estimated standard errors is acceptable if its absolute value is less than .10.

Analysis of variance (ANOVA) was used as descriptive partitioning of variance of the observed relative biases to determine the impacts of the three design factors (i.e., ICC, cluster size, and degree of cross-classification). Given that the purpose of using ANOVA in the present study is descriptive rather than inferential, the *p* value of the *F*-test was not reported. Instead, η^2 effect sizes were computed and reported as a measure of practical significance.

Results

The results indicated that the parameter estimates of the correctly specified model accurately reflected the population parameter values. Table 5 presents the means of the relative biases of the estimated variance components (i.e., $\hat{\theta}$, $\hat{\psi}$, $\hat{\tau}$) of the misspecified model. The means of the estimated variance components themselves are in the parentheses. The results showed that for the misspecified model (i.e., without considering F2 in the model), there was no bias in the estimated level-1 residual variance ($\hat{\theta}$) under all conditions, which indicated that ignoring the crossed factor at the third level did not affect the estimated variance of the first level. The estimated level-2 variance ($\hat{\psi}$) and the estimated variance of the remaining crossed factor ($\hat{\tau}$) were positively biased to some degree.

Bias in the Estimated Level-2 Variance ($\hat{\psi}$)

Compared to the population value of .2, the estimated level-2 variance had a positive bias ranging from .220 to .735, which indicated that a large amount of the variance of F2 was redistributed to level 2. The source of the bias was then investigated by conducting an ANOVA with the three design factors (i.e., ICC, cluster size, and degree of cross-classification). Table 6 presents the effect sizes of the main and the interaction effects, with the average bias at each level of the design factors[3]. For the observed bias in $\hat{\psi}$, the intraclass correlation had the largest effect η^2 = .520, which means that more than 50% of the variation of the bias in $\hat{\psi}$ could be explained by the

[3] Effect sizes less than .0001 were not reported.

factor of intraclass correlation. This is a big effect considering that all of the effects together could not exceed 100%. As expected, the larger the ICC of the ignored crossed random factor (i.e., F2), the greater the bias was found in the estimated level-2 variance.

The cluster size only had a very small effect on the bias ($\eta^2 = .001$). The bias in $\hat{\psi}$ was always greater when the ignored crossed factor had a smaller cluster size. The degree of cross-classification also had a very small effect on the bias ($\eta^2 = .002$). Larger bias was found with larger degree of cross-classification.

Bias in the Estimated Variance of the Remaining Crossed Factor ($\hat{\tau}$)

The bias in the estimated variance of the remaining crossed factor F1 ($\hat{\tau}$) was small, ranging from -.031 to .160 (see Table 5). Positive bias that were greater than .10 occurred under conditions in which the intraclass correlation coefficient of F2 was larger (i.e., $ICC_{F2} = .15$) and the degree of cross-classification was 50%.

The ANOVA results indicated that the intraclass correlation had a very small effect on the bias ($\eta^2 = .005$). Table 6 shows that as the ICC of F2 increased, the bias in $\hat{\tau}$ also increased. The degree of cross-classification also had a very small effect ($\eta^2 = .004$). The bias increased slightly as the degree of cross-classification decreased from 100% to 50%. The cluster size did not have a statistically significant effect and the effect size is less than .001. One possible reason could be that the differences between the levels in this design factor were not large enough to have impact on the observed bias.

TABLE 5

Relative Biases of the Estimated Variance Components of the Misspecified Model

(Simulation One)

ICC	Cluster Size	%Cross	B($\hat{\theta}$)	B($\hat{\psi}$)	B($\hat{\tau}$)
$ICC_{F1} = .15$ & $ICC_{F2} = .05$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	.002 (.601)	.235 (.247)	.007 (.151)
		50%	-.002 (.599)	.240 (.248)	-.031 (.146)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	.000 (.600)	.245 (.249)	.007 (.151)
		50%	.000 (.600)	.220 (.244)	.000 (.150)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	.003 (.602)	.230 (.246)	.000 (.150)
		50%	.000 (.600)	.220 (.244)	.007 (.151)
$ICC_{F1} = .10$ & $ICC_{F2} = .10$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	.002 (.601)	.480 (.296)	.010 (.101)
		50%	.000 (.600)	.450 (.290)	.030 (.103)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	.002 (.601)	.480 (.296)	.010 (.101)
		50%	.000 (.600)	.450 (.290)	.030 (.103)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	.002 (.601)	.465 (.293)	-.010 (.099)
		50%	.002 (.601)	.440 (.288)	.060 (.106)
$ICC_{F1} = .05$ & $ICC_{F2} = .15$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	.000 (.600)	.735 (.347)	.000 (.050)
		50%	.002 (.601)	.725 (.345)	.100 (.055)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	.002 (.601)	.725 (.345)	.000 (.050)
		50%	.002 (.601)	.700 (.340)	.160 (.058)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	.000 (.600)	.715 (.343)	.020 (.051)
		50%	.002 (.601)	.665 (.333)	.120 (.056)

Note. Figures in the parentheses are parameter estimates.

TABLE 6

Effects of Design Factors on the Relative Biases (Simulation One)

Design Factors	Random		Level-1 Covariate		Level-2 Covariate		Level-3 Covariate	
	$\hat{\psi}$	$\hat{\tau}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$
ICC								
η^2	.520	.005	.466	N/A[a]	.443	.556	.377	N/A
ICC_{F1}=.15 & ICC_{F2}=.05	.23	.00	-.13	N/A	-.11	.06	-.07	N/A
ICC_{F1}=.10 & ICC_{F2}=.10	.46	.02	-.25	N/A	-.22	.13	-.16	N/A
ICC_{F1}=.05 & ICC_{F2}=.15	.71	.06	-.40	N/A	-.34	.19	-.27	N/A
Cluster Size								
η^2	.001	.0005	.334	N/A	.354	.007	.315	N/A
n_{F1}=25 & n_{F2}=10	.48	.02	-.15	N/A	-.12	.12	-.08	N/A
n_{F1}=20& n_{F2}=20	.47	.04	-.26	N/A	-.23	.13	-.16	N/A
n_{F1}=10 & n_{F2}=25	.46	.03	-.38	N/A	-.32	.14	-.26	N/A
%Cross								
η^2	.002	.004	.001	N/A	.001	.001	.002	N/A
100%	.48	.05	-.27	N/A	-.23	.13	-.17	N/A
50%	.46	.06	-.26	N/A	-.22	.12	-.16	N/A
ICC*Cluster Size								
η^2	.0003	.0002	.0007	N/A	.0002	.0004	.0003	N/A
ICC*%Cross								
η^2	.0002	.0005	.0004	N/A	.0002	--[b]	.0006	N/A
Cluster Size*% Cross								
η^2	.0002	.0004	.00001	N/A	--	--	--	N/A
ICC*Cluster Size*%Cross								
η^2	.0002	.0006	.0001	N/A	--	.0002	--	N/A

[a] "N/A" indicates that the estimates were unbiased.
[b] "--" indicates that the η^2 was less than .0001.

Bias in the Estimated Fixed Effects

No matter which level the covariate was associated with, the biases in the estimated intercept $\hat{\gamma}_0$ and the estimated regression coefficient $\hat{\gamma}_1$ were acceptable under the misspecified model, which indicated that ignoring a crossed factor did not affect the estimation of the fixed effects themselves.

Biases in the Standard Errors of the Fixed Effects with a Level-1 Covariate

Table 7 presents the means of biases of the estimated standard errors of the fixed effects (i.e., $\hat{S}_{\hat{\gamma}_0}$ and $\hat{S}_{\hat{\gamma}_1}$) with a covariate of level-1, level-2, and the remaining level-3 crossed factor (i.e., F1). The results showed that the bias in $\hat{S}_{\hat{\gamma}_1}$ was acceptable under the misspecified model. However, the estimated standard error associated with the intercept (i.e., $\hat{S}_{\hat{\gamma}_0}$) had a moderate to large negative bias, ranging from -.051 to -.552.

The ANOVA results indicated that the intraclass correlation accounted for almost half of the variation of the observed bias in $\hat{S}_{\hat{\gamma}_0}$ (η^2 = .466). Table 6 shows that the bias became larger as the ICC of F2 increased. The cluster size explained about one third of the variation of the observed bias (η^2 = .334). The larger the cluster size, the larger the bias was. The degree of cross-classification only had a very small effect on the bias (η^2 = .001). There was slightly more bias when the data was fully cross-classified.

Biases in the Standard Errors of the Fixed Effects with a Level-2 Covariate

The estimated standard error associated with the intercept (i.e., $\hat{S}_{\hat{\gamma}_0}$) was negatively biased under the misspecified model, ranging from -.043 to -.460 (see Table 7). The ANOVA results indicated that the biggest effect was for the intraclass correlation (η^2 = .443). The larger the ICC of the ignored crossed factor, the greater the bias was (see Table 6). The cluster size had the second largest effect on the bias (η^2 = .354). The bias became larger as the cluster size of F2 increased. The degree of cross-classification only had a very small effect on the bias (η^2 = .001). There was slightly more bias when the data was fully cross-classified.

The estimated standard error associated with the regression coefficient (i.e., $\hat{S}_{\hat{\gamma}_1}$) had a small to moderate positive bias under the misspecified model, ranging from .06 to .20 (see Table 7). The ANOVA results indicated that the intraclass correlation had the largest effect on the bias ($\eta^2 = .556$). The larger the ICC of F2, the larger the bias was (see Table 6). The cluster size had a very small effect on the bias ($\eta^2 = .007$). Larger cluster size was related with larger bias. The degree of cross-classification also had a very small effect on the bias ($\eta^2 = .001$). There was slightly more bias when the data was fully cross-classified.

Biases in the Estimated Standard Errors of the Fixed Effects with a Covariate of the Remaining Crossed Factor

The estimated standard error associated with the intercept (i.e., $\hat{S}_{\hat{\gamma}_0}$) was negatively biased under the misspecified model, ranging from -.021 to -.398 (see Table 7). The ANOVA results indicated that the intraclass correlation and the cluster size were two major design factors that influenced the bias (Intraclass correlation: $\eta^2 = .377$; Cluster Size: $\eta^2 = .315$). Table 6 shows that the larger the ICC and the cluster size of F2, the greater the bias. The degree of cross-classification only had a very small effect on the bias ($\eta^2 = .002$). The bias was slightly larger when the data was fully cross-classified.

The relative bias of the estimated standard error associated with the regression coefficient (i.e., $\hat{S}_{\hat{\gamma}_1}$) in the misspecified model was acceptable under all conditions (see Table 7). However, there was a trend that as the intraclass correlation of F2 increased, the relative bias became larger.

TABLE 7

Relative Biases of the Estimated Standard Errors of the Fixed Effects of the Misspecified Model (Simulation One)

ICC	Cluster Size	%Cross	Level-1 covariate		Level-2 covariate		Level-3 covariate	
			$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$
$ICC_{F1} = .15$ & $ICC_{F2} = .05$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	-.052 (.090)	.000 (.005)	-.052 (.093)	.061 (.052)	-.021 (.127)	.010 (.182)
		50%	-.051 (.088)	.000 (.005)	-.043 (.094)	.061 (.052)	-.032 (. 130)	.000 (.186)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	-.122 (.081)	.000 (.049)	-.111 (.085)	.061 (.052)	-.074 (.115)	.000 (.163)
		50%	-.112 (.081)	.000 (.005)	-.110 (.085)	.061 (.052)	-.063 (.116)	.012 (.167)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	-.231 (.060)	.000 (.005)	-.189 (.066)	.062 (.053)	-.134 (.086)	.024 (.122)
		50%	-.223 (.061)	.000 (.005)	-.178 (.067)	.062 (.053)	-.121 (.087)	.012 (.122)
$ICC_{F1} = .10$ & $ICC_{F2} = .10$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	-.143 (.075)	.000 (.005)	-.113 (.080)	.100 (.055)	-.073 (.107)	.014 (.154)
		50%	-.121 (.076)	.000 (.005)	-.111 (.080)	.100 (.055)	-.074 (.109)	.018 (.156)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	-.262 (.069)	.000 (.005)	-.223 (.074)	.120 (.055)	-.158 (.098)	.024 (.140)
		50%	-.251 (.069)	.000 (.005)	-.223 (.075)	.120 (.055)	-.139 (.099)	.029 (.142)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	-.398 (.052)	.000 (.005)	-.342 (.059)	.122 (.056)	-.262 (.075)	.032 (.106)
		50%	-.372 (.054)	.000 (.005)	-.331 (.060)	.122 (.056)	-.253 (.076)	.052 (.107)
$ICC_{F1} = .05$ & $ICC_{F2} = .15$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	-.272 (.057)	.000 (.005)	-.210 (.064)	.183 (.059)	-.142 (.083)	.031 (.119)
		50%	-.236 (.059)	.000 (.005)	-.210 (.065)	.184 (.059)	-.134 (.084)	.035 (.120)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	-.423 (.053)	.000 (.005)	-.283 (.061)	.198 (.059)	-.272 (.077)	.054 (.109)
		50%	-.398 (.056)	.000 (.005)	-.354 (.062)	.183 (.058)	-.254 (.080)	.083 (.114)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	-.552 (.043)	.000 (.005)	-.460 (.052)	.203 (.059)	-.398 (.062)	.071 (.087)
		50%	-.529 (.044)	.000 (.005)	-.460 (.053)	.203 (.059)	-.384 (.064)	.094 (.089)

Note. Figures in the parentheses are estimated standard errors.

Discussion

Consistent with findings of previous studies (Fielding, 2002; Meyers, 2004; Meyers & Beretvas, 2006; Raudenbush & Bryk, 2002), ignoring a random crossed factor does not affect the estimates of the fixed effects themselves, including the intercept and the regression coefficient associated with covariates at any level. However, ignoring a random crossed factor causes biases in the standard errors of the fixed effects and the estimated variance components.

When a crossed factor at the highest level (i.e., level-3) is ignored, its variance component is split into *two* parts. One part (generally the larger part) is redistributed to level-2, causing a moderate to large positive bias in the estimated variance of the level-2 factor B. The magnitude of the bias is greatly affected by the intraclass correlation of the ignored crossed factor (i.e., F2) and slightly affected by the cluster size of this ignored factor and the degree of cross-classification between the two crossed factors. As expected, the larger the intraclass correlation, that is, the larger the variance of F2, the greater the bias is found in the estimated variance of the level-2 factor B. Larger cluster size and partial cross-classification is related to smaller bias.

The other part (generally the smaller part) of the variance of the ignored factor is redistributed to the remaining level-3 crossed factor (i.e., F1), causing a small to moderate positive bias in the estimated variance of F1. The magnitude of the bias is related to the intraclass correlation of the ignored level-3 factor (i.e., F2) and the degree of cross-classification between the two crossed factors. Interestingly, under all conditions, the level-1 residual variance is not biased, which means that the variance of F2 is not redistributed to the level-1 residual variance.

An important finding is that the standard error of the regression coefficient could be either unbiased or overestimated, depending on the level with which the covariate is associated. If the covariate is associated with the remaining level-3 crossed factor F1, the bias in the estimated standard error of the regression coefficient is acceptable under all the conditions in this study. If the covariate is at level-2, the estimated standard error of its regression coefficient is positively biased (i.e., overestimated), with a small to moderate magnitude. If the covariate is at level-1, the standard error of its regression coefficient is unbiased.

No matter which level the covariate belongs to, ignoring the level-3 crossed factor, F2, causes moderate to large negative bias in the estimated standard error of the intercept. This bias increases as the intraclass correlation and cluster size of F2 become larger.

Simulation Two

Methods

Data Generation

In this study, the cross-classification occurred at the intermediate level, that is, the second level (see Figure 3). A continuous covariate (x) associated with the level-1 unit A, or the level-2 crossed factor F1, or the level-3 factor B is included in the model as follows[4]:

$$y = \gamma_0 + \gamma_1 x + \mu + \nu + \omega + \varepsilon \tag{3.6}$$

where γ_0 and γ_1 are the fixed intercept and regression coefficient, respectively. μ is the random effect of the level-2 crossed factor F1 ($\mu \sim N(0,\psi)$), ν is the random effect of the other level-2 crossed factor F2 ($\nu \sim N(0,\tau)$), ω is the random effect of the level-3 factor B ($\omega \sim N(0,\varsigma)$), and ε is the level-1 residual ($\varepsilon \sim N(0,\theta)$).

As in Simulation One, γ_0 and γ_1 were set to be .1 and .5, respectively. ς and θ were set to be .1 and .5 respectively, and the conditional variance of y was set to be 1.0 (i.e., $Var(y \mid x) = 1.0$). The magnitude of ψ and τ varied under different conditions.

There were 12500 level-1 units A in the dataset. They were cross-classified by level-2 random factors F1 and F2. The cluster sizes of F1 and F2 (i.e., n_{F1} and n_{F2}) varied under different conditions. Both F1 and F2 were nested within level-3 factor B. There were 25 clusters in factor B, which means that there were 500 level-1 units within each cluster in factor B for the balanced design.

Design Factors

As in Simulation One, three design factors were manipulated: (1) the conditional intraclass correlation coefficients ($ICC_{F1} = .30$ and $ICC_{F2} = .10$, $ICC_{F1} = ICC_{F2} = .20$, $ICC_{F1} = .10$ and $ICC_{F2} = .30$), (2) the cluster sizes ($n_{F1} = 25$ and $n_{F2} = 10$,

[4] For the same reasons as stated in Simulation One, covariates associated with the ignored crossed factor (i.e., level-2 factor F2) were not considered in the model.

$n_{F1} = n_{F2} = 20$, $n_{F1} = 10$ and $n_{F2} = 25$), and (3) the degree of cross-classification ($\%Cross = 100\%$ and $\%Cross = 50\%$).

Similar to Simulation One, Simulation Two involved a total of 18 conditions (i.e., 3 ICCs × 3 Cluster sizes × 2 Degrees of cross-classification). For each condition, 500 datasets were generated and each data set was then analyzed using two different models: (1) the correctly specified model (i.e., cross-classified model), in which both F1 and F2 were included as crossed factors; (2) the misspecified model (i.e., hierarchical linear model), in which one of the level-2 crossed factors (i.e., F2) was excluded from the model.

Analysis

Relative biases of the parameter estimates and the corresponding standard error estimates were calculated using formula (3.4) and (3.5), respectively. ANOVA was used to determine the contribution of the three design factors to the observed bias.

Results

Similar to Simulation One, the results of Simulation Two showed that the parameter estimates of the correct model accurately reflected the population values. Table 8 presents the means of the relative biases of the estimated variance components (i.e., $\hat{\theta}$, $\hat{\psi}$, $\hat{\varsigma}$) of the misspecified model. The numbers in the parentheses are the means of the estimated variance components themselves.

Bias in the Estimated Level-1 Residual Variance ($\hat{\theta}$)

Compared to the population value of .6, the estimated level-1 residual variance ($\hat{\theta}$) under the misspecified model was overestimated. Table 8 shows that the bias ranged from .182 to .588, which indicated that a large amount of variance of F2 was redistributed to level 1.

Table 9 presents the effect sizes of the main and interaction effects of the three design factors and the average bias at each level of the design factors[5]. The factor of intraclass correlation uniquely explained 95% of the variation in the bias of $\hat{\theta}$ (η^2 = .950). As expected, the larger the ICC of the ignored crossed factor, the greater the bias.

The cluster size only had a very small effect on the bias (η^2 = .002). Smaller cluster size was related to greater bias. The degree of cross- classification also had a very small effect on the bias (η^2 = .002). The bias was greater when the two factors were 100% crossed.

Bias in the Estimated Variance of the Remaining Crossed Factor ($\hat{\psi}$)

The bias in the estimated variance of the remaining crossed factor F2 ($\hat{\psi}$) ranged from -.005 to .150 (see Table 8). The relative biases in $\hat{\psi}$ were acceptable when the intraclass correlation coefficient of the ignored crossed factor (i.e., F2) was either .10 or .20. When ICC_{F2} increased to .30 and the two factors were partially crossed (i.e., *%Cross* = 50%), substantial amount of bias in $\hat{\psi}$ was found ($B(\hat{\psi}) = .150$).

The ANOVA results showed that the degree of cross-classification explained 10% of the variation of the bias (η^2 = .100). Table 9 shows that the bias was greater when the two factors were 50% crossed. The factor of intraclass correlation explained 6% of the variation of the bias (η^2 = .060). The larger the ICC of F2, the greater the bias.

[5] Effect sizes less than .0001 were not reported.

TABLE 8

Relative Biases of the Estimated Variance Components of the Misspecified Model

(Simulation Two)

ICC	Cluster Size	%Cross	B($\hat{\theta}$)	B($\hat{\psi}$)	B($\hat{\varsigma}$)
$ICC_{F1} = .30$ & $ICC_{F2} = .10$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	.196 (.598)	.006 (.302)	.040 (.104)
		50%	.190 (.595)	.013 (.304)	.040 (.104)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	.192 (.596)	.000 (.300)	.020 (.102)
		50%	.182 (.591)	.013 (.304)	.030 (.103)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	.188 (.594)	-.003 (.299)	.040 (.104)
		50%	.180 (.590)	.013 (.304)	.020 (.102)
$ICC_{F1} = .20$ & $ICC_{F2} = .20$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	.392 (.696)	-.005 (.199)	.020 (.102)
		50%	.382 (.691)	.015 (.203)	.030 (.103)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	.384 (.692)	-.005 (.199)	.090 (.109)
		50%	.364 (.682)	.045 (.209)	.070 (.107)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	.380 (.690)	-.005 (.199)	.080 (.108)
		50%	.360 (.680)	.050 (.210)	.080 (.108)
$ICC_{F1} = .10$ & $ICC_{F2} = .30$	$n_{F1} = 25$ & $n_{F2} = 10$	100%	.588 (.794)	.000 (.100)	.070 (.107)
		50%	.576 (.788)	.070 (.107)	.040 (.104)
	$n_{F1} = 20$ & $n_{F2} = 20$	100%	.576 (.788)	.000 (.100)	.120 (.112)
		50%	.552 (.776)	.120 (.112)	.090 (.109)
	$n_{F1} = 10$ & $n_{F2} = 25$	100%	.572 (.786)	.000 (.100)	.160 (.116)
		50%	.540 (.770)	.150 (.115)	.150 (.115)

Note. Figures in the parentheses are parameter estimates.

TABLE 9

Effects of Design Factors on the Relative Biases (Simulation Two)

Design Factors	Random			Level-1 Covariate		Level-2 Covariate		Level-3 Covariate	
	$\hat{\theta}$	$\hat{\psi}$	$\hat{\varsigma}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$
ICC									
η^2	.950	.060	.030	N/A[a]	.950	N/A	N/A	N/A	N/A
ICC_{F1}=.10 & ICC_{F2}=.30	.19	.01	.03	N/A	.07	N/A	N/A	N/A	N/A
ICC_{F1}=.20 & ICC_{F2}=.20	.38	.02	.06	N/A	.14	N/A	N/A	N/A	N/A
ICC_{F1}=.30 & ICC_{F2}=.10	.57	.03	.11	N/A	.21	N/A	N/A	N/A	N/A
Cluster Size									
η^2	.002	.009	.003	N/A	.010	N/A	N/A	N/A	N/A
n_{F1}=25 & n_{F2}=10	.39	.03	.04	N/A	.13	N/A	N/A	N/A	N/A
n_{F1}=20& n_{F2}=20	.38	.03	.07	N/A	.15	N/A	N/A	N/A	N/A
n_{F1}=10 & n_{F2}=25	.37	.04	.09	N/A	.16	N/A	N/A	N/A	N/A
%Cross									
η^2	.002	.100	.0002	N/A	.002	N/A	N/A	N/A	N/A
100%	.39	.00	.07	N/A	.15	N/A	N/A	N/A	N/A
50%	.37	.06	.06	N/A	.14	N/A	N/A	N/A	N/A
ICC*Cluster Size									
η^2	.0004	.0001	.0003	N/A	.0003	N/A	N/A	N/A	N/A
ICC*%Cross									
η^2	.0004	.0006	.0002	N/A	.0002	N/A	N/A	N/A	N/A
Cluster Size*%Cross									
η^2	.0002	.0001	--[b]	N/A	.0003	N/A	N/A	N/A	N/A
ICC*Cluster Size*%Cross									
η^2	--	.0005	.0002	N/A	--	N/A	N/A	N/A	N/A

[a] "N/A" indicates that the estimates were unbiased.
[b] "--" indicates that the η^2 was less than .0001.

The cluster size only explained less than 1% of the variation of the bias (η^2 = .009). The larger the cluster size, the larger the bias.

Bias in the Estimated Level-3 Variance ($\hat{\varsigma}$)

The bias in the estimated level-3 variance ($\hat{\varsigma}$) ranged from .020 to .160 (see Table 8). When the intraclass correlation of F2 was .10, there was almost no bias in $\hat{\varsigma}$. The largest bias ($B(\hat{\varsigma}) = .160$) occurred under the condition with large intraclass correlation (i.e., $ICC_{F2} = .30$) , large cluster size of the ignored level-2 factor F2 (i.e., $n_{F2} = 25$), and full cross-classification (i.e., *%Cross* = 100% between F1 and F2).

The ANOVA results indicated that the intraclass correlation had a small effect (η^2 = .030). Table 9 shows that the larger the ICC of F2, the greater the bias. The cluster size only had a very small effect on the bias (η^2 = .003). As the cluster size of F2 increased, the bias increased. The degree of cross-classification was statistically nonsignificant and the effect size was less than .001.

Bias in the Estimated Fixed Effects

Consistent with the findings of Simulation One, no matter which level the covariate was associated with, the estimated intercept $\hat{\gamma}_0$ and regression coefficient $\hat{\gamma}_1$ were unbiased under the misspecified model, which confirmed the statement that ignoring a crossed factor did not affect the estimation of the fixed effects themselves.

Biases in the Estimated Standard Errors of the Fixed Effects with a Level-1 Covariate

Table 10 presents the means of biases of the estimated standard errors of the fixed effects (i.e., $\hat{S}_{\hat{\gamma}_0}$ and $\hat{S}_{\hat{\gamma}_1}$) with a covariate of level-1, the remaining level-2 crossed factor (i.e., F1), and level-3. The results showed that the only bias was in the estimated standard error associated with the regression coefficient (i.e., $\hat{S}_{\hat{\gamma}_1}$) when the covariate was at level-1. The bias was positive, ranging from .052 to .233.

The ANOVA results indicated that the factor of intraclass correlation explained 95% of the variation of the bias (η^2 = .950). Table 9 shows that the greater the intraclass correlation of the ignored crossed factor, the greater the bias. The cluster size only

TABLE 10

Relative Biases of the Estimated Standard Errors of the Fixed Effects of the Misspecified Model (Simulation Two)

ICC	Cluster Size	%Cross	Level-1 covariate		Level-2 covariate		Level-3 covariate	
			$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$	$\hat{S}_{\hat{\gamma}_0}$	$\hat{S}_{\hat{\gamma}_1}$
$ICC_{F1}=.30$ & $ICC_{F2}=.10$	$n_{F1}=25$ & $n_{F2}=10$	100%	.000 (.107)	.052 (.022)	.000 (.073)	.000 (.052)	.000 (.098)	.000 (.140)
		50%	.000 (.069)	.081 (.014)	.000 (.073)	.000 (.052)	.000 (.097)	.012 (.139)
	$n_{F1}=20$ & $n_{F2}=20$	100%	.000 (.067)	.081 (.014)	.000 (.072)	.000 (.047)	.000 (.099)	.000 (.140)
		50%	.000 (.068)	.081 (.014)	.000 (.072)	.000 (.047)	.000 (.097)	.000 (.139)
	$n_{F1}=10$ & $n_{F2}=25$	100%	.021 (.067)	.081 (.014)	.000 (.068)	.000 (.034)	.000 (.095)	.000 (.135)
		50%	.000 (.066)	.081 (.014)	.000 (.070)	.033 (.035)	.000 (.097)	.000 (.137)
$ICC_{F1}=.20$ & $ICC_{F2}=.20$	$n_{F1}=25$ & $n_{F2}=10$	100%	.000 (.067)	.150 (.015)	.000 (.071)	.022 (.044)	.000 (.096)	.000 (.137)
		50%	.000 (.067)	.150 (.015)	.000 (.070)	.022 (.044)	.012 (.097)	.000 (.139)
	$n_{F1}=20$ & $n_{F2}=20$	100%	.012 (.069)	.150 (.015)	.000 (.070)	.034 (.040)	.000 (.098)	.000 (.139)
		50%	.000 (.068)	.150 (.015)	.000 (.071)	.034 (.040)	.000 (.097)	.012 (.138)
	$n_{F1}=10$ & $n_{F2}=25$	100%	.000 (.067)	.150 (.015)	.000 (.069)	.031 (.030)	.013 (.096)	.000 (.135)
		50%	.000 (.067)	.150 (.015)	.000 (.069)	.031 (.030)	.000 (.098)	.000 (.139)
$ICC_{F1}=.10$ & $ICC_{F2}=.30$	$n_{F1}=25$ & $n_{F2}=10$	100%	.000 (.067)	.233 (.016)	.000 (.068)	.032 (.033)	.000 (.095)	.000 (.135)
		50%	.000 (.066)	.233 (.016)	.000 (.068)	.063 (.034)	.000 (.095)	.013 (.136)
	$n_{F1}=20$ & $n_{F2}=20$	100%	.000 (.068)	.233 (.016)	.000 (.069)	.072 (.031)	.000 (.096)	.000 (.138)
		50%	.013 (.068)	.233 (.016)	.012 (.070)	.073 (.032)	.000 (.158)	.000 (.223)
	$n_{F1}=10$ & $n_{F2}=25$	100%	.000 (.069)	.233 (.016)	.000 (.069)	.042 (.024)	.000 (.098)	.000 (.139)
		50%	.009 (.069)	.233 (.016)	.012 (.070)	.000 (.023)	.000 (.096)	.012 (.139)

Note. Figures in the parentheses are parameter estimates.

explained 1% of the variation of the bias (η^2 = .010). The larger the cluster size, the larger the bias. The degree of cross-classification between F1 and F2 only had a very small effect on the bias (η^2= .002). The bias was slightly larger when the two factors were fully crossed.

Discussion

When a crossed factor at the intermediate level (i.e., level-2) is ignored, its variance component is split into *three* parts (i.e., one large part and two small parts). One part (generally the large part) is redistributed to level-1, causing a moderate to large positive bias in the estimated level-1 residual variance. The magnitude of the bias is greatly affected by the intraclass correlation of this ignored level-2 crossed factor. The larger the intraclass correlation, the greater the bias. The cluster size of F2 and the degree of cross-classification has small impact on the bias of the level-1 residual variance. Larger cluster size and partial cross-classification are related to smaller bias in the estimation of the level-1 residual variance. Another part (generally one of the two small parts) of the variance of the ignored level-2 crossed factor (i.e., F2) is redistributed to level-3, causing a small to moderate positive bias of the estimated level-3 variance. Larger bias is related to larger intraclass correlation and larger cluster size of the ignored level-2 crossed factor.

The third part (the other small part) of the variance in F2 is redistributed to the remaining level-2 crossed factor (i.e., F1), causing a small to moderate positive bias of the estimated variance of F1. Larger intraclass correlation and larger cluster size of the ignored level-2 crossed factor (i.e., F2) together with partial cross-classification of the two crossed factors are related to larger bias in the estimated variance of F1.

Similar to the findings in Simulation One, the bias of the estimated standard error of the regression coefficient depends on the level with which the covariate is associated. If the covariate is at level-1, the standard error of its regression coefficient is positively biased, with a small to moderate magnitude. The bias increases as the ICC and the cluster size of the ignored level-2 crossed factor increases. If the covariate is associated with either the remaining level-2 crossed factor or the level-3 factor, its standard error will not be biased under all conditions. Interestingly, no matter which level the covariate belongs to, the estimated standard error of the intercept is unbiased.

CHAPTER IV

STUDY TWO: IGNORING CROSS-CLASSIFIED STRUCTURES IN LONGITUDINAL DATA

In longitudinal data, the cross-classified structure is a result of individuals' change of cluster membership over time. If individuals do not move to different clusters over time, the data structure would be strictly hierarchical with repeated measures nested within individuals and individuals nested within higher level clusters, such as classes and schools. As long as one individual moves to a different cluster over time, the hierarchy would be destroyed and the data becomes cross-classified with repeated measures cross-classified by individuals and clusters.

In educational research cross-classified longitudinal data are very common because students often move to different classes or schools. However, because of the complexity of cross-classified model, researchers either exclude participants who move to different schools from their sample of analysis to keep the hierarchical structure of their data (De Fraine, Van Landeghem, Van Damme, & Onghena, 2005; McCoach, O'Connel, Reis, & Levitt, 2006), or simply ignore the cross-classified structure of the data and use hierarchical linear models (George & Thomas, 2000; Ma & Ma, 2004).

In this study, the impact of ignoring the cross-classified structures in longitudinal data was investigated, and factors that could affect the biases of parameter estimates were examined.

Methods

Data Generation

For concreteness, consider the example in which student's math scores are measured for 4 times, with one-year time interval between two measures. All students are in a closed system of *K* schools. For simplicity, students are only allowed to switch schools at the second time point. The model used to generate the data is a two-level cross-classified model as follows:

$$\text{Level-1:} \quad y_{tjk} = \pi_{0jk} + \pi_{1jk} x_{tjk} + \varepsilon_{tjk}$$

$$\text{Level-2:} \quad \pi_{0jk} = \gamma_{00} + \gamma_{01} w_{jk} + \gamma_{02} z_k + \mu_{00j} + \nu_{00k}$$

$$\pi_{1jk} = \gamma_{10} + \gamma_{11} w_{jk} + \gamma_{12} z_k + \mu_{10j} \qquad (5.1)$$

where t indexes the occasions (t = 1, 2, 3, 4), j indexes students (j = 1, .., J), and k indexes schools (k = 1, .., K).

At level 1 a time variable x_{tjk} measuring the number of years between occasion t and the initial occasion was generated. For each student, x_{tjk} took on value of 0, 1, 2, and 3 corresponding to occasion 1, 2 ,3 and 4, respectively. y_{tjk} is the math score of student j in school k at occasion t. π_{0jk} is the expected level of math score at the initial occasion for student j. π_{1jk} is the annual rate of growth in math for student j. ε_{tjk} is a random within-subject residual that is normally distributed with mean 0 and variance θ.

For each student a time-invariant covariate w_{jk} (e.g., SES) with the standard normal distribution was generated. For each school a covariate z_k (e.g., teacher student ratio) with the standard normal distribution was generated. In the level-2 model w_{jk} and z_k were used to predict π_{0jk} and π_{1jk}. μ_{00j} is the deviation of student *j* from the overall initial level of math achievement; and μ_{10j} is the deviation of student *j* from the overall growth rate of math achievement. The joint distribution of μ_{00j} and μ_{10j} is bivariate normal with mean 0 and variance and covariance $\boldsymbol{\psi} = \begin{bmatrix} \psi_{00} & \\ \psi_{01} & \psi_{11} \end{bmatrix}$. ν_{00k} is the residual school random effect, that is, an expected deflection of the growth curve associated with studying in school *k*, conditional on w_{jk} and z_k. ν_{00k} is assumed to be normally distributed with mean 0 and variance τ.

In this simulation, γ_{00} was set to be .10 and γ_{01}, γ_{02}, γ_{10}, γ_{11}, and γ_{12} were set to be .50 for all conditions. θ (i.e., within-subject residual variance) was set to be .4. The magnitude of $\boldsymbol{\psi}$ (i.e., the variance and covariance matrix of the two student random effects μ_{00j} and μ_{10j}) and τ (i.e., the variance of the school random effect ν_{00k}) varied according to different conditions.

Design Factors

Five design factors were manipulated in the study. Three of them represent student-level effects: mobility rate, number of students per school, and variance and covariance of student random effects. Two of them represent school-level effects: number of schools, and variance of school random effects.

Combining the five factors, this study involves 48 conditions. For each condition, 200 datasets were generated. Each dataset was analyzed using 2 models: (1) the correct model, that is, the cross-classified model used to generate the data; (2) a misspecified model, in which students' mobility is ignored. The misspecified model is actually a three level hierarchical linear model assuming that repeated measures are nested within students and students are nested within the schools that they belong to at time 1. The 3-level hierarchical linear model is specified as

Level-1: $y_{tjk} = \pi_{0jk} + \pi_{1jk} x_{tjk} + \varepsilon_{tjk}$

Level-2: $\pi_{0jk} = \theta_{0k} + \gamma_{01} w_{jk} + \mu_{00j}$

$\pi_{1jk} = \theta_{1k} + \gamma_{11} w_{jk} + \mu_{10j}$

Level-3: $\theta_{0k} = \gamma_{00} + \gamma_{02} z_k + \nu_{00k}$

$$\theta_{1k} = \gamma_{10} + \gamma_{12} z_k . \quad (5.2)$$

The data were generated using SAS 9.1 (SAS Institute Inc., 2004) and the models were estimated using the **lmer** function in the R package **lme4** (Bates & Sarkar, 2007). According to Bates (2005), the **lme4** package is "fast and reliable" on examples that are "currently considered typical multilevel modeling examples" (p. 16). A comparison of **lmer** and SAS PROC MIXED showed that **lmer** was about 9 times faster than SAS PROC MIXED in analyzing the cross-classified data generated in the present study.

Mobility Rate

Mobility rate is defined as the proportion of students moving to different schools each year. According to the statistics provided by Texas Education Agency, the average mobility rate in Texas was 21% in 2004-2005 and the standard deviation is about 15%. Accordingly, three levels of mobility rate were chosen, with a low mobility rate of 5%, a medium rate of 20%, and a large rate of 35%. For simplicity, students were randomly selected to move to any other schools in the system only once at time 2.

Number of Schools

A review of 41 published studies using longitudinal multilevel data from 2004 to 2005 in the ERIC database showed that the number of schools varied greatly across studies, ranging from less than 10 schools to more than 3000 schools and the distribution was highly positively skewed. Two levels were selected in this factor, that is, 25 vs. 50 schools, which represent small and medium number of schools, respectively.

Number of Students per School

According to the review, the number of students per school ranged from 10 to 425, with a mean of 84 and a standard deviation of 90. The two levels in this factor were chosen to be 50 and 100 students per school to represent small and medium school size, respectively.

Variance of School Random Effect v_{00k} (τ)

Following previous simulation studies (Moerbeek, 2004; Meyers & Beretvas, 2006), this factor consists of two levels: (1) a small size $\tau = .1$ and (2) a medium size $\tau = .2$.

Variance and Covariance of Student Random Effect μ_{00j} *and* μ_{10j} ($\boldsymbol{\psi}$)

There were two levels in this factor: (1) a medium size $\boldsymbol{\psi} = \begin{bmatrix} .20 & \\ .05 & .10 \end{bmatrix}$ according to the criteria provided by Raudenbush and Liu (2001), and (2) a small size $\boldsymbol{\psi} = \begin{bmatrix} .10 & \\ .025 & .05 \end{bmatrix}$ that is half the magnitude of the medium size.

Analysis

The parameter estimates from the correct model and the misspecified model were summarized across the 200 iterations for each condition and compared with the other model's estimates (where relevant) and with known population values. The relative bias of parameter estimates is calculated using formula (3.4) and the relative bias of estimated standard errors is calculated using formula (3.5). Analysis of variance (ANOVA) was conducted to determine which of the study's factors contributed to the observed bias.

Results

Random Effects

In both the correct model and the misspecified model, there are five parameters related to the random effects. They are the variance of the school random effects (τ), variance of the student random effects associated with the initial level of math achievement (ψ_{00}), variance of the student random effects associated with the growth

TABLE 11

Relative Biases of the Estimated Variance Components and the Standard Errors of the Fixed Effects of the Misspecified Model

Mobility Rate	N Schools	N Student	τ	Ψ	$\hat{\tau}$	$\hat{\psi}_{11}$	$\hat{\psi}_{01}$	$\hat{S}_{\hat{\gamma}_{00}}$	$\hat{S}_{\hat{\gamma}_{02}}$
5%	25	50	.1	small	-.050 (.095)	.020 (.051)	.038 (.026)	-.045 (.064)	-.275 (.050)
5%	25	50	.1	medium	-.039 (.096)	.000 (.100)	.020 (.051)	-.029 (.066)	-.246 (.052)
5%	25	50	.2	small	-.050 (.190)	.020 (.051)	.080 (.027)	-.033 (.089)	-.411 (.056)
5%	25	50	.2	medium	-.046 (.191)	.010 (.101)	.020 (.051)	-.033 (.089)	-.351 (.061)
5%	25	100	.1	small	-.067 (.093)	.020 (.051)	.080 (.027)	-.045 (.063)	-.464 (.037)
5%	25	100	.1	medium	-.060 (.094)	.010 (.101)	.020 (.051)	-.031 (.063)	-.582 (.041)
5%	25	100	.2	small	-.081 (.184)	.020 (.051)	.120 (.028)	-.055 (.086)	-.574 (.040)
5%	25	100	.2	medium	-.064 (.187)	.010 (.101)	.040 (.052)	-.043 (.088)	-.526 (.046)
5%	50	50	.1	small	-.040 (.096)	.020 (.051)	.040 (.026)	-.021 (.046)	-.277 (.034)
5%	50	50	.1	medium	-.040 (.096)	.010 (.101)	.020 (.051)	-.021 (.046)	-.229 (.037)
5%	50	50	.2	small	-.045 (.191)	.020 (.051)	.080 (.027)	-.031 (.063)	-.409 (.039)
5%	50	50	.2	medium	-.036 (.193)	.010 (.101)	.040 (.052)	-.016 (.063)	-.338 (.043)
5%	50	100	.1	small	-.060 (.094)	.020 (.051)	.040 (.026)	-.043 (.044)	-.435 (.026)
5%	50	100	.1	medium	-.040 (.096)	.000 (.100)	.040 (.052)	-.028 (.045)	-.391 (.028)
5%	50	100	.2	small	-.060 (.188)	.020 (.051)	.080 (.027)	-.031 (.062)	-.569 (.028)
5%	50	100	.2	medium	-.054 (.189)	.010 (.101)	.060 (.053)	-.046 (.062)	-.523 (.031)
20%	25	50	.1	small	-.224 (.078)	.060 (.053)	.160 (.029)	-.121 (.058)	-.552 (.030)
20%	25	50	.1	medium	-.186 (.081)	.030 (.103)	.306 (.065)	-.090 (.061)	-.514 (.034)
20%	25	50	.2	small	-.219 (.156)	.102 (.055)	.360 (.034)	-.132 (.079)	-.656 (.033)
20%	25	50	.2	medium	-.186 (.163)	.051 (.105)	.180 (.059)	-.108 (.083)	-.625 (.036)

Note. Figures in the parentheses are parameter estimates.

TABLE 11 (Continued)

Mobility Rate	N Schools	N Student	τ	$\boldsymbol{\psi}$	$\hat{\tau}$	$\hat{\psi}_{11}$	$\hat{\psi}_{01}$	$\hat{S}_{\hat{\gamma}_{00}}$	$\hat{S}_{\hat{\gamma}_{02}}$
20%	25	100	.1	small	-.233 (.077)	.040 (.052)	.160 (.029)	-.136 (.057)	-.667 (.023)
20%	25	100	.1	medium	-.183 (.082)	.020 (.102)	.080 (.054)	-.091 (.060)	-.629 (.026)
20%	25	100	.2	small	-.228 (.154)	.100 (.055)	.360 (.034)	-.141 (.079)	-.750 (.024)
20%	25	100	.2	medium	-.188 (.162)	.040 (.104)	.180 (.059)	-.121 (.080)	-.687 (.026)
20%	50	50	.1	small	-.212 (.079)	.040 (.052)	.160 (.029)	-.108 (.041)	-.552 (.021)
20%	50	50	.1	medium	-.170 (.083)	.030 (.103)	.080 (.054)	-.085 (.043)	-.500 (.024)
20%	50	50	.2	small	-.204 (.159)	.100 (.055)	.360 (.034)	-.108 (.058)	-.657 (.023)
20%	50	50	.2	medium	-.172 (.166)	.050 (.105)	.200 (.060)	-.091 (.060)	-.627 (.025)
20%	50	100	.1	small	-.204 (.080)	.040 (.052)	.200 (.030)	-.111 (.040)	-.674 (.015)
20%	50	100	.1	medium	-.167 (.083)	.020 (.102)	.080 (.054)	-.106 (.042)	-.638 (.017)
20%	50	100	.2	small	-.206 (.159)	.100 (.055)	.360 (.034)	-.109 (.057)	-.754 (.016)
20%	50	100	.2	medium	-.174 (.165)	.050 (.105)	.180 (.059)	-.108 (.058)	-.723 (.018)
35%	25	50	.1	small	-.343 (.066)	.060 (.053)	.280 (.032)	-.194 (.054)	-.632 (.025)
35%	25	50	.1	medium	-.299 (.070)	.040 (.104)	.180 (.059)	-.152 (.056)	-.594 (.028)
35%	25	50	.2	small	-.337 (.133)	.140 (.057)	.520 (.038)	-.187 (.074)	-.713 (.027)
35%	25	50	.2	medium	-.282 (.144)	.079 (.108)	.280 (.064)	-.161 (.078)	-.698 (.029)
35%	25	100	.1	small	-.353 (.065)	.080 (.054)	.280 (.032)	-.212 (.052)	-.721 (.019)
35%	25	100	.1	medium	-.296 (.070)	.040 (.104)	.140 (.057)	-.169 (.054)	-.697 (.020)
35%	25	100	.2	small	-.348 (.130)	.160 (.058)	.600 (.040)	-.208 (.073)	-.799 (.019)
35%	25	100	.2	medium	-.291 (.142)	.081 (.108)	.280 (.064)	-.174 (.075)	-.776 (.021)
35%	50	50	.1	small	-.333 (.067)	.078 (.054)	.280 (.032)	-.176 (.039)	-.629 (.018)
35%	50	50	.1	medium	-.267 (.073)	.040 (.104)	.140 (.057)	-.138 (.041)	-.594 (.020)

Note. Figures in the parentheses are parameter estimates.

TABLE 11 (Continued)

Mobility Rate	N Schools	N Student	τ	Ψ	$\hat{\tau}$	$\hat{\psi}_{11}$	$\hat{\psi}_{01}$	$\hat{S}_{\hat{\gamma}_{00}}$	$\hat{S}_{\hat{\gamma}_{02}}$
35%	50	50	.2	small	-.320 (.136)	.160 (.058)	.560 (.039)	-.175 (.053)	-.716 (.019)
35%	50	50	.2	medium	-.269 (.146)	.080 (.108)	.280 (.064)	-.145 (.056)	-.692 (.021)
35%	50	100	.1	small	-.340 (.066)	.080 (.054)	.280 (.032)	-.186 (.037)	-.727 (.013)
35%	50	100	.1	medium	-.283 (.072)	.040 (.104)	.140 (.057)	-.151 (.039)	-.697 (.014)
35%	50	100	.2	small	-.327 (.135)	.160 (.058)	.560 (.039)	-.185 (.052)	-.795 (.013)
35%	50	100	.2	medium	-.274 (.145)	.080 (.108)	.300 (.065)	-.152 (.054)	-.776 (.015)

Note. Figures in the parentheses are parameter estimates.

rate of math achievement (ψ_{11}), covariance of the two student random effects (ψ_{01}), and the within-student residual variance (θ). In the correct model, the parameter estimates accurately reflected the population values. In the misspecified model, the estimates of τ, ψ_{11}, and ψ_{01} were biased. Table 11 presents the mean relative biases of the three parameter estimates. The numbers in the parentheses are the parameter estimates themselves.

Bias in the Estimated Variance of the School Random Effects ($\hat{\tau}$)

In general, the variance of school random effects was underestimated. When the mobility rate was 5%, the bias was small and acceptable. As the mobility rate increased to 20% and 35%, the bias became larger, ranging from -.167 to -.353.

Table 12 presents the contributions of the five design factors to the observed bias and the average bias at each level of the design factors[6]. For the observed bias in $\hat{\tau}$, the mobility rate explained about 20% of its variation (η^2 = .194). As the mobility rate increased, the bias became greater. The magnitude of the variance and covariance matrix of the student random effects ($\boldsymbol{\psi}$) had a very small effect (η^2 = .005). The smaller the $\boldsymbol{\psi}$ matrix, the greater the bias in $\hat{\tau}$.

There was a very small interaction effect between the mobility rate and the magnitude of the $\boldsymbol{\psi}$ matrix (i.e., variance and covariance of student random effects) (η^2 = .002). As the mobility increased, the increase in the bias was greater when the $\boldsymbol{\psi}$ was small (see Figure 4).

Bias in the Estimated Variance of the Student Random Effects Associated with the Growth Rate ($\hat{\psi}_{11}$)

When the mobility rate was 5%, the bias in $\hat{\psi}_{11}$ was acceptable. When the mobility rate increased to 20%, small positive bias showed up under conditions of medium variance of school random effects ($\tau = .2$) and small $\boldsymbol{\psi}$ matrix (i.e., variance and covariance of student random effects). When the mobility rate increased to 35%, the positive bias became apparent under all conditions (see Table 11).

[6] Effect sizes less than .0001 were not reported.

TABLE 12

Effects of Design Factors on the Relative Biases (Study Two)

		$\hat{\tau}$	$\hat{\psi}_{11}$	$\hat{\psi}_{01}$	$\hat{S}_{\hat{\gamma}_{00}}$	$\hat{S}_{\hat{\gamma}_{02}}$
Main Effect						
Mobility Rate (MR)	η^2	.194	.081	.180	.800	.661
	5%	-.049	.008	.051	-.034	-.396
	20%	-.192	.037	.200	-.110	-.636
	35%	-.313	.064	.314	-.172	-.700
N of Schools (N_SCH)	η^2	.0005	-- [a]	.0003	.012	--
	25	-.190	--	.184	-.112	--
	50	-.179	--	.193	-.098	--
N of Students (N_STU)	η^2	.0002	.0001	.001	.007	.142
	50	-.181	.035	.181	-.100	-.517
	100	-.188	.037	.196	-.110	-.638
τ	η^2	.0002	.019	.071	.001	.099
	.1	-.181	.025	.120	-.103	-.527
	.2	-.188	.048	.256	-.107	-.628
$\boldsymbol{\psi}$	η^2	.005	.166	.061	.027	.015
	small	-.203	.069	.251	-.115	-.597
	medium	-.167	.003	.125	-.095	-.557
Interaction Effect						
MR* τ	η^2	.0003	.006	.025	.0002	.002
MR* $\boldsymbol{\psi}$	η^2	.002	.067	.018	.008	.001
τ * $\boldsymbol{\psi}$	η^2	--	.015	.008	--	--
MR*N_SCH	η^2	.0002	--	.0001	.0005	--
MR* N_STU	η^2	.0002	.0002	--	.0004	.010
N_SCH*N_STU	η^2	--	.0001	.0002	.0002	--
N_SCH* τ	η^2	--	.0001	--	.0001	--
N_STU* τ	η^2	--	.0002	--	--	.0003
N_STU* $\boldsymbol{\psi}$	η^2	--	.0002	--	.0001	--
MR*N_SCH* τ	η^2	--	.0002	.0002	--	--
MR*N_SCH* $\boldsymbol{\psi}$	η^2	.0002	--	--	--	--
MR*N_STU* τ	η^2	.0003	--	--	--	.0005
MR*N_STU* $\boldsymbol{\psi}$	η^2	.0001	.0002	--	--	--
MR* τ * $\boldsymbol{\psi}$	η^2	.0003	.004	.003	--	.0001

[a] "--" indicates that the η^2 was less than .0001.

Table 12 (Continued)

		$\hat{\tau}$	$\hat{\psi}_{11}$	$\hat{\psi}_{01}$	$\hat{S}_{\hat{\gamma}_{00}}$	$\hat{S}_{\hat{\gamma}_{02}}$
N_STU* τ * $\boldsymbol{\psi}$	η^2	.0001	-- [a]	--	--	--
N_SCH*N_STU* $\boldsymbol{\psi}$	η^2	.0001	.0002	--	--	--
MR*N_STU* τ * $\boldsymbol{\psi}$	η^2	.0005	--	--	--	--
MR*N_SCH*N_STU* τ	η^2	--	--	.0003	--	--
MR*N_SCH* τ * $\boldsymbol{\psi}$	η^2	--	.0002	.0002	--	--
MR*N_SCH*N_STU* τ * $\boldsymbol{\psi}$	η^2	.0002	.0001	--	--	--

[a] "--" indicates that the η^2 was less than .0001.

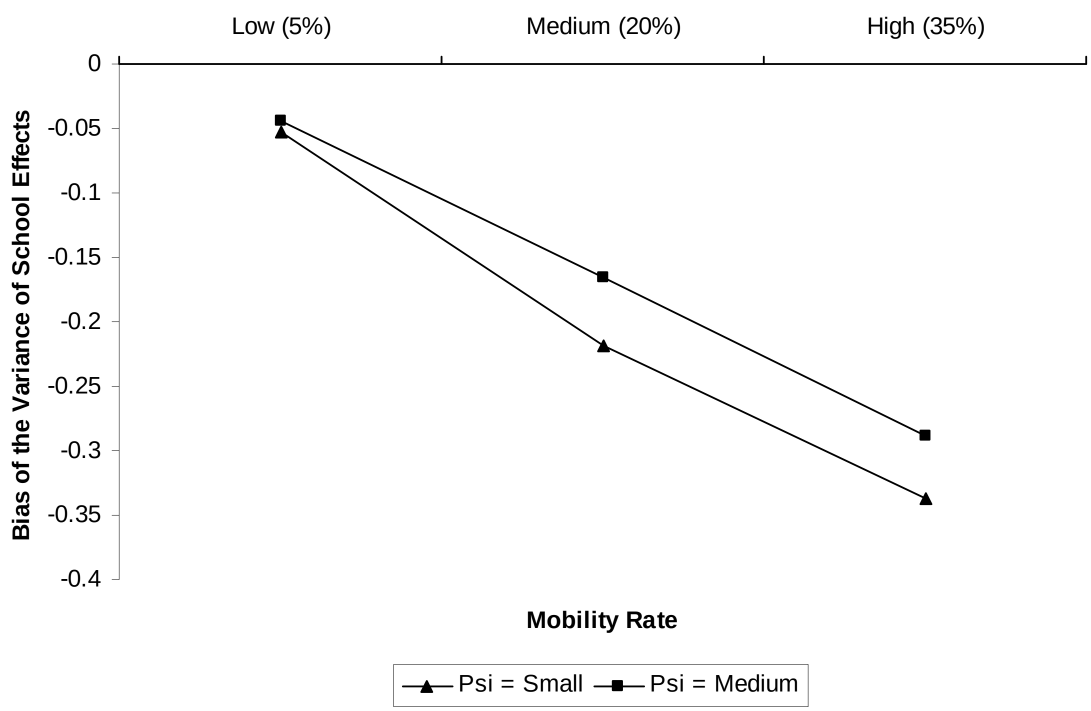

FIGURE 4 Effects of the mobility rate and $\boldsymbol{\psi}$ on the bias of $\hat{\tau}$

The ANOVA results indicated that the magnitude of the $\boldsymbol{\psi}$ matrix was the most important factor influencing the observed bias (η^2 = .166). The smaller the $\boldsymbol{\psi}$ matrix, the greater the bias. The effect of the mobility rate was about half that of the $\boldsymbol{\psi}$ matrix (η^2 = .081). Apparently, the increase in the mobility rate caused larger bias in $\hat{\psi}_{11}$. The magnitude of τ had a small effect on the bias (η^2 = .019). As τ increased, the bias became larger.

There were significant two-way interaction effects between the mobility rate, the magnitude of $\boldsymbol{\psi}$, and the magnitude of τ (Mobility rate*$\boldsymbol{\psi}$: η^2 = .067; Mobility rate*τ : η^2 = .006; $\boldsymbol{\psi}$ *τ : η^2 = .015) and a significant three-way interaction effect (η^2 = .004). Figure 5 shows that when the $\boldsymbol{\psi}$ matrix was large, there was almost no bias for all levels of mobility rate and τ. When the $\boldsymbol{\psi}$ matrix was small, as the mobility rate increased, the increase in the bias was greater with larger τ.

Bias in the Estimated Covariance of the Student Random Effects ($\hat{\psi}_{01}$)

When the mobility rate was 5%, the bias in $\hat{\psi}_{01}$ was small and acceptable. As the mobility rate increased to 20% and 35%, the positive bias became larger, ranging from .08 to .60.

The ANOVA results indicated that the mobility rate explained about 20% of the variation of the bias in $\hat{\psi}_{01}$ (η^2 = .180). The magnitude of τ had a small effect (η^2 = .071) with larger τ causing larger bias. The magnitude of the $\boldsymbol{\psi}$ matrix had a very small effect (η^2 = .001) with larger $\boldsymbol{\psi}$ causing smaller bias.

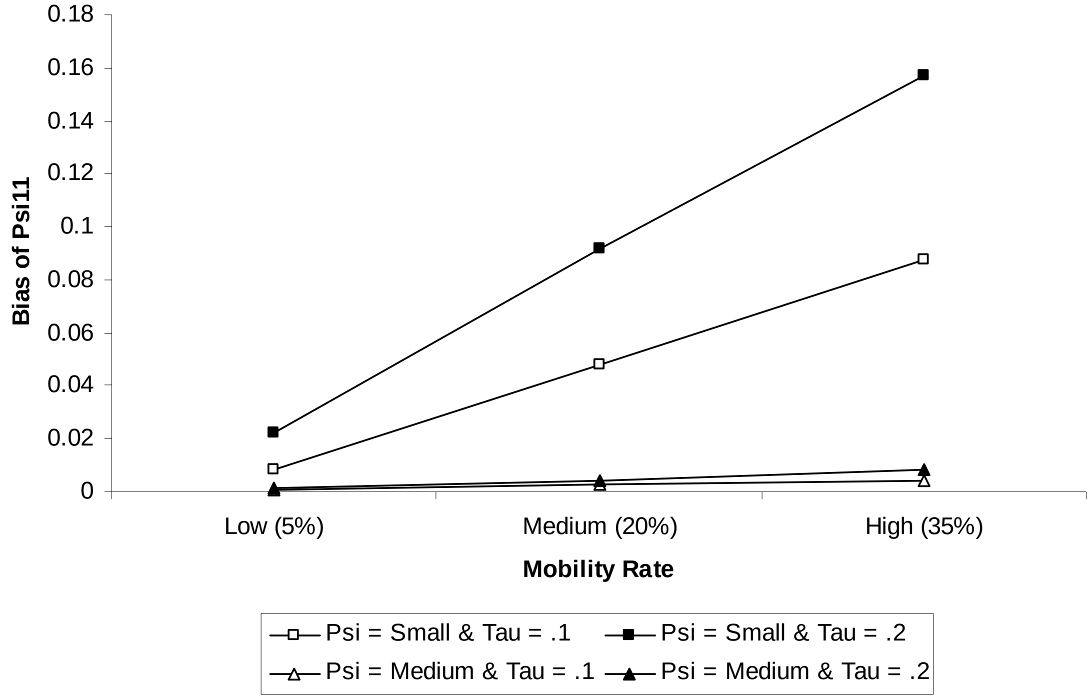

FIGURE 5 Effects of the mobility rate, τ , and $\boldsymbol{\psi}$ on the bias of $\hat{\psi}_{11}$

There were significant two-way interaction effects between the mobility rate, the magnitude of $\boldsymbol{\psi}$, and the magnitude of τ (Mobility rate* $\boldsymbol{\psi}$: $\eta^2 = .018$; Mobility rate* τ : $\eta^2 = .025$; $\boldsymbol{\psi} * \tau$: $\eta^2 = .008$) and a significant three-way interaction effect ($\eta^2 = .003$). Figure 6 shows that as the mobility rate increased, the increase of the bias was the greatest with larger τ and smaller $\boldsymbol{\psi}$. For conditions that have larger τ with larger $\boldsymbol{\psi}$ or smaller τ with smaller $\boldsymbol{\psi}$, the increase of the bias was similar. The increase of the bias was the smallest when smaller τ was combined with larger $\boldsymbol{\psi}$.

The number of students per school had a very small effect ($\eta^2 = .001$). The more students were there in each school, the greater bias was found in $\hat{\psi}_{01}$.

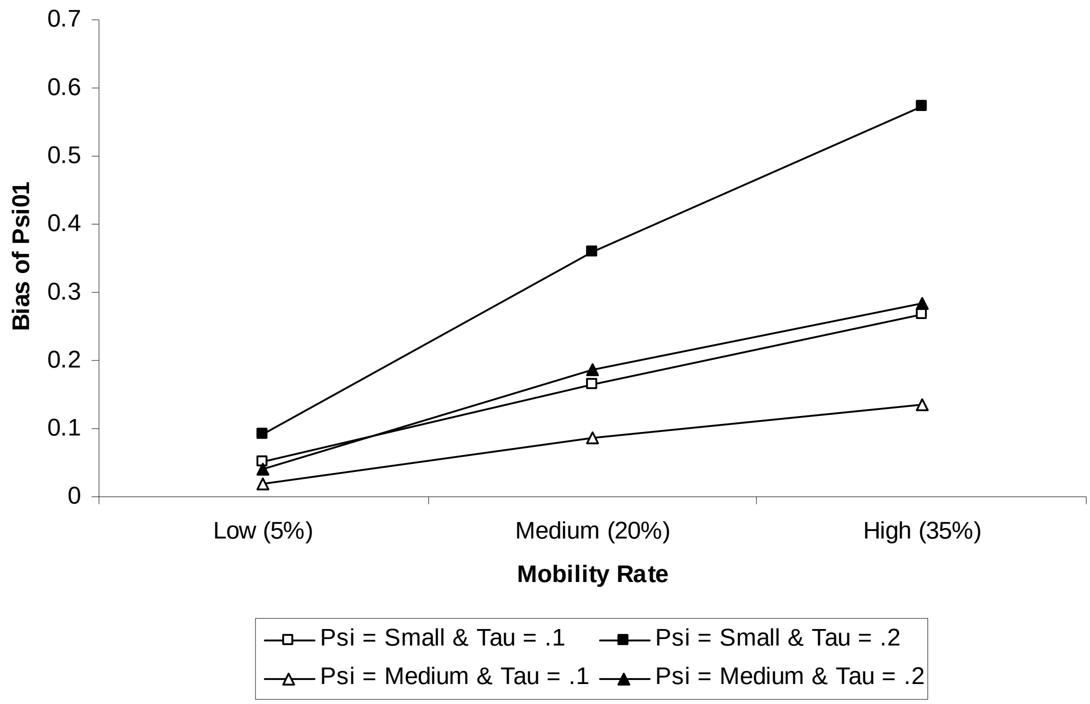

FIGURE 6 Effects of the mobility rate, τ , and $\boldsymbol{\psi}$ on the bias of $\hat{\psi}_{01}$

Fixed Effects and Standard Errors

Consistent with previous findings, the estimated fixed effects themselves were unbiased in the misspecified model. However, the standard error of the intercept and the standard error of the regression coefficient associated with the school-level predictor *z* were biased. The last two columns in Table 11 present the means of the relative biases of the estimated standard errors of the two fixed effects.

Bias in the Estimated Standard Error of the Intercept ($\hat{S}_{\hat{\gamma}_{00}}$)

When the mobility rate was 5%, the bias in the estimated standard error of the intercept ($\hat{S}_{\hat{\gamma}_{00}}$) was acceptable. When the mobility rate increased to 20% and 35%, the bias in $\hat{S}_{\hat{\gamma}_{00}}$ became larger, ranging from -.085 to -.212.

The ANOVA results showed that the mobility rate had the largest effect on the bias (η^2 = .800). As expected, the larger the mobility rate, the greater the bias in the estimated standard error of the intercept (see Table 12). The magnitude of the variance and covariance matrix of student random effect μ_{00j} and μ_{10j} ($\boldsymbol{\psi}$) had a small effect on the bias (η^2 = .027). The bias in $\hat{S}_{\hat{\gamma}_{00}}$ was greater when $\boldsymbol{\psi}$ was small. There was a very small interaction effect between the mobility rate and the magnitude of $\boldsymbol{\psi}$ (η^2 = .008). As the mobility rate increased, the increase in the bias was greater when the $\boldsymbol{\psi}$ was small (see Figure 7).

The number of schools also had a small effect (η^2 = .012). As the number of schools increased, the bias became smaller. The number of students per school had a very small effect (η^2 = .007). As the number of students per school increased, the bias became slightly greater. The magnitude of the variance of school random effects (τ) also had a very small effect (η^2 = .001). Larger τ was associated with greater bias.

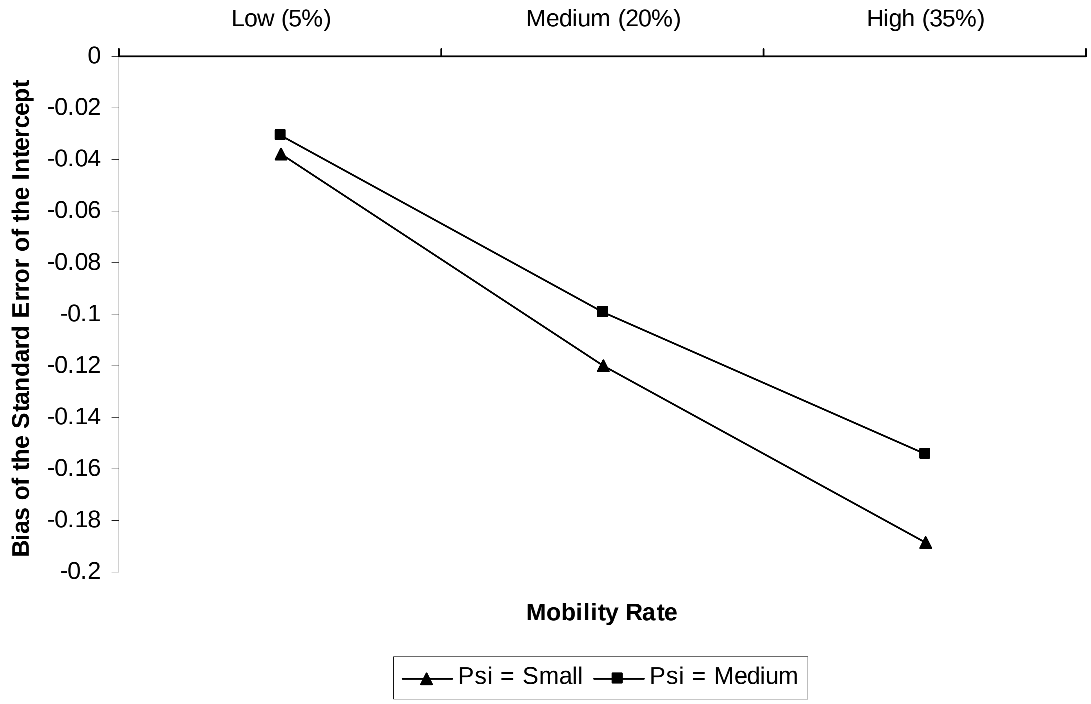

FIGURE 7 Effects of the mobility rate and ψ on the bias of $\hat{S}_{\hat{\gamma}_{00}}$

Bias in the Estimated Standard Error of the Regression Coefficient of the School-level Covariate ($\hat{S}_{\hat{\gamma}_{02}}$)

The standard error associated with the school-level covariate had a moderate to large negative bias under all conditions, ranging from -.246 to -.795. The mobility rate had the largest effect (η^2 = .661), which was almost 5 times that of the second largest effect. As expected, the larger the mobility rate, the greater the bias in $\hat{S}_{\hat{\gamma}_{02}}$. The number of student per school had the second largest effect (η^2 = .142). As the number of students per school increased, the bias became larger. There was a small interaction effect between the mobility rate and the number of students per school (η^2 = .010). As the mobility rate increased, the increase in the bias was slightly smaller when the number of students per school was large (see Figure 8).

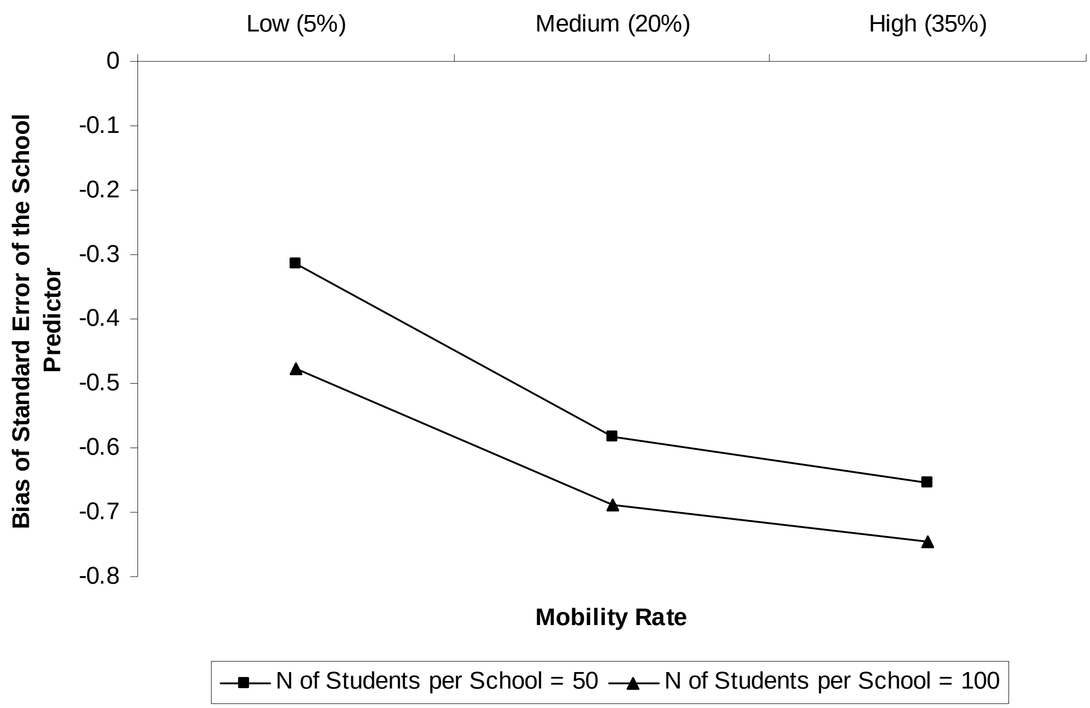

FIGURE 8 Effects of the mobility rate and the number of students per school on the bias of $\hat{S}_{\hat{\gamma}_{02}}$

The magnitude of the variance of school random effects (τ) had a small effect (η^2 = .099). Larger variance of school random effects was associated with greater bias. There was a very small interaction effect between the mobility rate and the variance of school random effects (η^2 = .002). As the mobility increased, the increase in the bias was slightly smaller with larger variance of school random effects (see Figure 9).

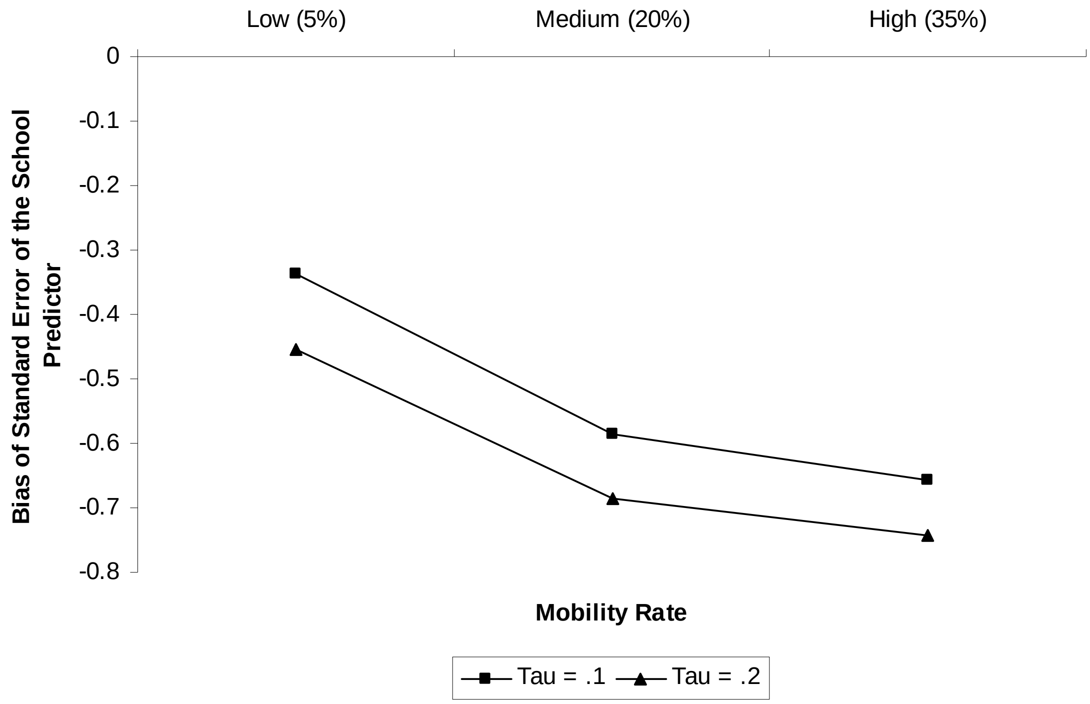

FIGURE 9 Effects of the mobility rate and τ on the bias of $\hat{S}_{\hat{\gamma}_{02}}$

The magnitude of $\boldsymbol{\psi}$ matrix (i.e., variance and covariance of student random effects) also had a small effect on the bias ($\eta^2 = .015$). The bias in $\hat{S}_{\hat{\gamma}_{02}}$ was greater when $\boldsymbol{\psi}$ was small. There was a very small interaction effect between mobility rate and the magnitude of $\boldsymbol{\psi}$ ($\eta^2 = .001$). As the mobility rate increased, the increase in the bias was greater when the $\boldsymbol{\psi}$ was large (see Figure 10).

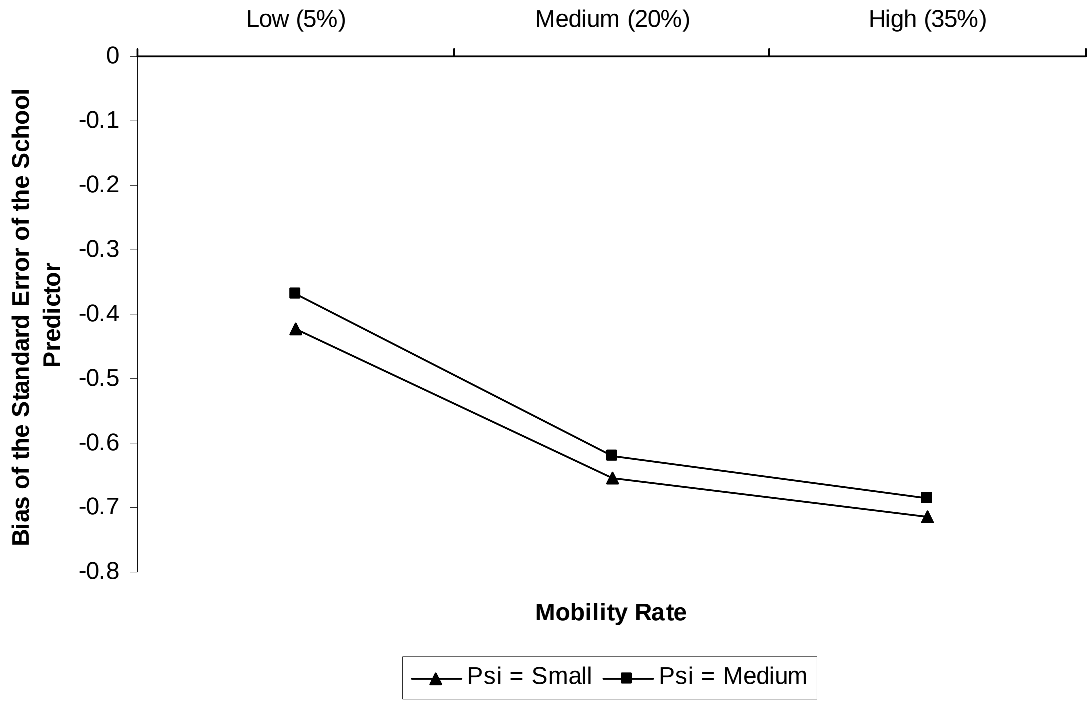

FIGURE 10 Effects of the mobility rate and $\boldsymbol{\psi}$ on the bias of $\hat{S}_{\hat{\gamma}_{02}}$

Discussion

In longitudinal data with cross-classified structures, such as repeated measures cross-classified by students and schools, when the school level is analyzed as a hierarchical level above the student level instead of a crossed random factor, part of the school level variance is added to the student level, causing underestimation of the school-level variance τ and overestimation of the student-level variance ψ_{11} and covariance ψ_{01}. Interestingly, the within-student variance θ and the between-student variance ψ_{00} are not affected.

Similar to cross-sectional data, in longitudinal data ignoring the cross-classified structures causes underestimation of the standard error of the intercept. Because the intercept of model (5.1) is related to the initial level of the outcome variable, underestimation of its standard error would cause inflated Type I error in testing whether the initial level of the outcome variable is different from zero.

Modeling the school level as the top level in a hierarchy rather than as a crossed factor also causes underestimated standard errors of the regression coefficient associated with the school-level predictors. However, the estimated standard errors of the regression coefficients associated with the student-level predictors are unbiased. The reason might be that students' mobility only destroys the nested structure between students and schools and repeated measures are still nested within students even if students move to different schools over time. The standard error of the regression coefficient associated with the time variable is unbiased either, because the estimated within-student variance is unbiased.

Mobility rate is a very important factor that affects the observed biases of those parameter estimates and standard error estimates. As expected, the greater the mobility rate, the larger the biases. With a low mobility rate (i.e., 5%), the biases of the parameter estimates and standard error estimates are all acceptably small (i.e., less than 5% for parameter estimates and less than 10% for standard error estimates), except for the biases of the estimated standard errors of the regression coefficients associated with school-level predictors. Accordingly, if the mobility rate is low and the school-level predictors are not of interest, researchers could ignore the cross-classification and use hierarchical linear models. When the mobility rate is moderate or high, based on the simulation results, it is recommended that researchers use cross-classified models.

The variance of school random effects (τ) and the variance and covariance of student random effects ($\boldsymbol{\psi}$) also affect the observed bias. It was found that greater τ and smaller $\boldsymbol{\psi}$ cause larger bias. It is recommended that researchers should consider using cross-classified models when there is large variance at the school level and relatively small variance at the student level, although this is rare in real data.

The number of students per school has a small effect on the bias of the standard error of the intercept and the standard errors of the regression coefficients of the school-level predictors. This result is reasonable because given a certain mobility rate, the larger number of students per school means the more students switching schools, which would cause larger bias of some estimates.

CHAPTER V

IMPLICATIONS AND CONCLUSIONS

In educational research, cross-classified multilevel data are often encountered. For example, in cross-sectional studies students are cross-classified by schools and neighborhoods; in longitudinal studies repeated measures are cross-classified by students and schools if students transfer schools over time. Cross-classified random effects models (CCREMs) provide flexible modeling techniques that are well matched to these cross-classified data structures. For example, in cross-sectional studies, CCREMs can be used to analyze effects of multiple contexts simultaneously (e.g., school, family, and neighborhood). For longitudinal studies CCREMs can be used to accommodate students' mobility and assess the effect of schools more precisely.

Despite the fact that CCREMs provide more modeling flexibility, major textbooks on multilevel modeling have chapters about CCREMs (e.g., Raudenbush & Bryk, 2002; Hox, 2002; Snijders & Bosker, 1999), and many multilevel modeling computer programs are capable of estimating CCREMs (e.g., HLM 6.0, MLwiN 2.0, SAS PROC MIXED, R package **lme4**), researchers in substantive areas seldom use the technique. In cross-sectional studies, researchers may only examine the effect of one higher-level factor, say schools, and ignore other potential crossed factors such as neighborhoods. In longitudinal studies, researchers may simply ignore the mobility of students and treat schools as a hierarchical level above students rather than a crossed factor.

It is very important for researchers to understand the functioning of CCREMs and the consequences of misspecifying CCREMs. However, little research has been conducted assessing when it is necessary to use CCREMs. The only introductory investigation was conducted by Meyers and his colleagues (Meyers, 2004; Meyers & Beretvas, 2006). It was found that if the cross-classified structure is ignored in a 2-level cross-classified data, the fixed effects estimates themselves are unaffected, however, the standard errors associated with the incorrectly modeled variables are underestimated; the bias of the standard errors gets worse as the variance attributable to the factor that was modeled incorrectly increases; and the level one variance and the variance of the remaining modeled crossed factor were overestimated.

The present study extended Meyers' study mainly in two directions. First, the present study investigated the impact of misspecifying CCREMs in more general 3-level cross-classified data, in which cross-classification could occur at either the top or the intermediate level. Second, the present study investigated the impact of misspecifying CCREMs in longitudinal data that have repeated measures cross-classified by students and schools.

The study had several important new findings. First, in cross-sectional cross-classified data the random effect estimates could be overestimated or unbiased when HLM is assumed instead of CCREM. Ignoring a crossed factor at the top level only causes overestimation of the variance of the remaining crossed factor and the intermediate level. Ignoring a crossed factor at the intermediate level causes overestimation of the variance of the remaining crossed factor, the top level, and the bottom level.

Second, in cross-sectional cross-classified data the standard errors of the fixed effects could be overestimated, unbiased, or underestimated. When a crossed factor at the highest level (i.e., level-3) is ignored, the standard errors corresponding to the covariates of either the intermediate level or the remaining crossed factor are overestimated. However, the standard errors corresponding to the covariates at the bottom level (i.e., level-1) are unbiased. When a crossed factor at the intermediate-level (i.e., level-2) is ignored, the standard errors corresponding to the covariates at the lowest level (i.e., level-1) are overestimated. However, the standard errors corresponding to the covariates at the intermediate and the highest level are unbiased. In general, because the overestimation of the standard errors leads to lower statistical power in testing the significance of regression coefficients, researchers should be aware that ignoring potential crossed random factors could lead to more conservative results.

On the other hand, if researchers are interested in the intercept, they should note that ignoring a crossed factor at the highest level causes the standard error of the intercept to be underestimated, which can lead to inflated Type I error rate.

Third, in longitudinal cross-classified data, ignoring the cross-classified structure and treating the school level as a hierarchical level above the student level causes underestimation of the school level variance and overestimation of the student level variance and covariance. The standard errors of the intercept and the regression coefficients associated with the school-level predictors are underestimated, which can lead to spurious significance for results.

The findings of the present study enhanced our understanding of the functioning of CCREMs and the importance of correctly modeling cross-classified multilevel data. Based on the simulation results, several suggestions are put forward for researchers to consider if the structure and the sample size of their data, and the models they consider to estimate are similar to those in the simulations. First, in cross-sectional studies researchers should assume CCREM instead of HLM if a crossed random factor is at the top level and its intraclass correlation is greater than .10 (see Table 6); or if a crossed factor is at the intermediate level and its intraclass correlation is greater than .20 (see Table 9). Second, in longitudinal data which have repeated measures cross-classified by individuals and clusters, researchers should find out the mobility rate at each time point and assume CCREM instead of HLM if the mobility rate is too high.

The present study only provided a preliminary investigation of the impact of misspecifying CCREMs in cross-classified longitudinal data. Consequently, there are several limitations. First, the present study assumed that the student's moving is completely random. In reality, however, students with certain characteristics (e.g., low SES) are more likely to move and students' choices of which school to move to are not random. In addition, it is common that students move to different schools more than once in the course of a study. The impact of ignoring the cross-classified structure in longitudinal multilevel data could be further investigated in future research by taking these issues into consideration.

REFERENCES

Bates, D. (2004). *Sparse matrix representations of linear mixed models.* Retrieved May 1, 2007, from http://www.stat.wisc.edu/%7Ebates/reports/MixedEffects.pdf

Bates, D. (2005). Fitting linear mixed models in R using the lme4 package. *R Newsletter, 5*, 27-30.

Bates, D. & Sarkar, D. (2007). *The lme4 package.* Retrieved May 14, 2007, from http://cran.r-project.org/doc/packages/lme4.pdf

De Fraine, B., Van Landeghem, G., Van Damme, J., & Onghena, P. (2005). An analysis of well-being in secondary school with multilevel growth curve models and multilevel multivariate models. *Quality and Quantity, 39,* 297-316.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components model. *Journal of the American Statistics Association, 76,* 341-353.

Fielding, A. (2002). Teaching groups as foci for evaluating performance in cost-effectiveness of GCE Advanced Level provision: Some practical methodological innovations. *School Effectiveness and School Improvement, 13*, 225-246.

George, R. & Thomas, G. (2000). Victimization among middle and high school students: A multilevel analysis. *The High School Journal, 84*, 48-57.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73*, 43-56.

Goldstein, H. (1995). *Multilevel statistical models.* London: Edward Arnold.

Hoogland, J.J. & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods and Research, 26,* 329-367.

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

Hutchison, D. & Healy, M. (2001). The effect of variance component estimates of ignoring a level in a multilevel model. *Multilevel Modeling Newsletter, 13,* 4-5.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute Inc.

Lockwood, J. R., Doran, H., & McCaffrey, D. F. (2003). Using R for estimating longitudinal student achievement models. *R Newsletter, 3*, 17-23.

Ma, X. & Ma, L. (2004). Modeling stability of growth between mathematics and science achievement during middle and high school. *Evaluation Review, 38*, 104-122.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*, 67-101.

McCoach, D. B., O'Connel, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology, 98*, 14-28.

Meyers, J. (2004). *The impact of the inappropriate modeling of cross-classified data structures.* Ph.D. dissertation, University of Texas at Austin. (UMI No. 3145342).

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

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39*, 129-149.

Opdenakker, M. C., & Van Damme, J. (2000). The importance of identifying levels in multilevel analysis: An illustration of the effects of ignoring the top or intermediate levels in school effectiveness research. *School Effectiveness and School Improvement, 11,* 103-130.

Rasbash J. and Browne W. J. (2006). Non-hierarchical multilevel models. In J. De Leeuw & I.G.G. Kreft (Eds.), *Handbook of quantitative multilevel analysis.* Retrieved May 12, 2006, from http://www.maths.nottingham.ac.uk/personal/pmzwjb/materials/nhmm.pdf

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*, 337-350.

Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2004). *A user's guide to MLwiN 2.0*. Bristol, UK: Center for Multilevel Modeling.

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

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

Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S. W. & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6,* 387-401.

SAS Institute Inc. (2004). *SAS/STAT 9.1 user's guide*. Cary, NC: SAS Institute Inc.

Snijders, T. A. B. & Bosker, R. J. (1999). Multilevel analysis: An introduction to basic and advanced multilevel modeling**.** London: SAGE.

Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models.* Boca Raton, FL: Chapman & Hall/CRC.

Van Landeghem, G., De Fraine, B., & Van Damme, J. (2005). The consequence of ignoring a level of nesting in multilevel analysis: A comment. *Multivariate Behavioral Research, 40,* 423-434.

Wolfinger, R. D., Tobias, R. D., & Sall, J. (1994). Computing Gaussian likelihoods and their derivatives for general linear mixed models. *SIAM Journal on Scientific Computing, 15*, 1294-1310.

APPENDIX A

SAS PROGRAM FOR CROSS-CLASSIFICATION AT THE TOP LEVEL

```
/* &N_F1 is the number of levels in F1; &V_F1 is the variance of F1;
&N_F2 is the number of levels in F2; &V_F2 is the variance of F2; &N_B
is the number of levels in B; &V_B is the variance of B; &N_A is the
number of level-1 units A; &V_A is the level-1 residual variance*/


Data F1_ran ;
 do F1_id = 1 to &N_F1;
  F1_ran=sqrt(&V_F1)*rannor(0);
 output;
 end;
Run;

Data F2_ran ;
 do F2_id = 1 to &N_F2;
  F2_ran=sqrt(&V_F2)*rannor(0);
 output;
 end;
Run;

Data B_ran ;
 do B_id= 1 to &N_B;
   B_ran=sqrt(&V_B)*rannor(0);
 output;
 end;
Run;

Data A_ran ;
 do A_id = 1 to &N_A;
  A_ran=sqrt(&V_A)*rannor(0);
 output;
 end;
Run;

/*IF X IS AT LEVEL 1, USE THIS*/
Data B_X;
 do B_id=1 to &N_B;
  do i=1 to 8;
   X=ranbin(0, 1, 0.5);
  output;
  end;
 output;
 end;
run;

/*IF X IS AT LEVEL 2, USE THIS*/
Data B_ran ;
 do B_id= 1 to &N_B;
   x=ranbin(0, 1, 0.5);
```

```
 output;
 end;
Run;

/*IF X IS AT LEVEL 3 USE THIS*/
Data F1_ran ;
 do F1_id = 1 to &N_F1;
   X=ranbin(0, 1, 0.5);
  output;
 end;
Run;

/*IF 100% CROSS-CLASSIFICATION, USE THIS*/
Data B_F1_F2;
 do B_id = 1 to &N_B;
  F1_id =INT(RANUNI(0)* &N_F1+1);
  F2_id =INT(RANUNI(0)* &N_F2+1);
 output;
end;
run;

/*IF 50% CROSS-CLASSIFICATION, USE THIS*/
Data B_F1_F2;
       do B_id = 1 to &N_B;
             if B_id<=(&N_B/2) then do;
             F1_id=INT(RANUNI(0)*(&N_F1/2)+1);
             F2_id=INT(RANUNI(0)*(&N_F2/2)+1);
             end;
             else do;
             F1_id =INT(RANUNI(0)* (&N_F1/2)+ (&N_F1/2)+1);
             F2_id =INT(RANUNI(0)* (&N_F2/2)+ (&N_F2/2)+1);
             end;
        output;
        end;
run;

Proc sort data=B_F1_F2;
 by F1_id;
Run;

Data B_F1ran_F2;
 merge B_F1_F2 F1_ran;
 by F1_id;
Run;

Proc sort data=B_F1ran_F2;
 by F2_id;
Run;

Data B_F1ran_F2ran;
 merge B_F1ran_F2 F2_ran;
 by F2_id;
Run;
```

```
Proc sort data=B_F1ran_F2ran;
 by B_id;
run;

Data Bran_F1ran_F2ran_X;
 merge B_ran B_F1ran_F2ran B_X ;
 by B_id;
run;

Data all;
 merge A_ran Bran_F1ran_F2ran_X;
 y= 0.1 + 0.5*X+F1_ran + F2_ran + B_ran + A_ran;
Run;

Proc mixed data=all;
class B_id F1_id F2_id;
model y= X/solution ;
random intercept / subject=F1_id;
random intercept / subject=F2_id;
random intercept / subject=B_id;
ODS OUTPUT SolutionF=F_F1_F2 ;
ods output covparms=R_F1_F2;
run;

Proc mixed data=all;
class B_id F1_id;
model y=X/solution;
random intercept / subject=F1_id;
random intercept / subject=F2_id;
ODS OUTPUT SolutionF=F_F1 ;
ods output covparms=R_F1;
run;

Data F_F1_int (rename=(estimate=int_F1 stderr=std_i_F1) );
set F_F1 ;
if effect='Intercept';
drop effect df probt tvalue;
run;

Data F_F1_X (rename=(estimate=X_F1 stderr=std_X_F1));
set F_F1;
if effect='X';
drop effect df probt tvalue;
run;

Data F_F1_F2_int (rename=(estimate=int_F12 stderr=std_i_F12) );
set F_F1_F2 ;
if effect='Intercept';
drop effect df probt tvalue;
run;

Data F_F1_F2_X (rename=(estimate=X_F12 stderr=std_X_F12) );
set F_F1_F2 ;
if effect='X';
```

```
drop effect df probt tvalue;
run;

data Fixed_1;
merge F_F1_F2_int F_F1_F2_X F_F1_int F_F1_X;
run;

Proc append out=Fixed force;
run;

Data R_F1_F1 (rename=(estimate=VF1_F1 stderr=std_VF1_F1));
set R_F1;
if covparm="Intercept" & subject="F1_id";
drop covparm subject zvalue probz;
run;

Data R_B_F1 (rename=(estimate=VB_F1 stderr=std_VB_F1));
set R_F1;
if covparm="Intercept" & subject="B_id";
drop covparm subject zvalue probz;
run;

Data R_A_F1 (rename=(estimate=VA_F1 stderr=std_VA_F1));
set R_F1;
if covparm="Residual" ;
drop covparm subject zvalue probz;
run;

Data R_F2_F12 (rename=(estimate=VF2_F12 stderr=std_VF2_F12));
set R_F1_F2;
if covparm="Intercept" & subject="F2_id";
drop covparm subject zvalue probz;
run;
Data R_F1_F12 (rename=(estimate=VF1_F12 stderr=std_VF1_F12));
set R_F1_F2;
if covparm="Intercept" & subject="F1_id";
drop covparm subject zvalue probz;
run;

Data R_B_F12 (rename=(estimate=VB_F12 stderr=std_VB_F12));
set R_F1_F2;
if covparm="Intercept" & subject="B_id";
drop covparm subject zvalue probz;
run;

Data R_A_F12 (rename=(estimate=VA_F12 stderr=std_VA_F12));
set R_F1_F2;
if covparm="Residual" ;
drop covparm subject zvalue probz;
run;

Data Random_1;
merge R_F2_F12 R_F1_F12 R_B_F12 R_A_F12
      R_F1_F1 R_B_F1 R_A_F1  ;
```

```
run;

Proc append out=Random force;
run;
```

APPENDIX B

SAS PROGRAM FOR CROSS-CLASSIFICATION AT THE INTERMEDIATE LEVEL

```
Data B_ran;
 do B_id = 1 to &N_B;
 B_ran=SQRT(&V_B)*rannor(0);
 output;
 end;
run;

/*IF 100% CROSS-CLASSIFIED, USE THIS*/
Data B_F1_F2_A_stack;
  do B_id=1 to &N_B;
     set B_ran;
     do i = 1 to &N_A;
       j =INT(RANUNI(0)*&N_F1+1);
       k =INT(RANUNI(0)*&N_F2+1);
       output;
     end;
 end;
run;

/*IF 50% CROSS-CLASSIFIED, USE THIS*/
Data B_F1_F2_A_stack;
 do B_id=1 to &N_B;
     set B_ran;
        do i = 1 to &N_A;
             if i<=(&N_A/2) then do;
             j=INT(RANUNI(0)* )*(&N_F1/2)+1);
             k =INT(RANUNI(0)* (&N_F2/2)+1);
             end;
             else do;
              j =INT(RANUNI(0)*(&N_F1/2)+ (&N_F1/2)+1);
              k =INT(RANUNI(0)* (&N_F2/2)+ (&N_F2/2)+1);
             end;
            output;
     end;
     end;
run;


Data B_F1_F2_A_total (drop=i j k index);
 set B_F1_F2_A_stack;
 do index=1 to &N_B;
 if B_id=index then do;
  A_id=i+&N_A*(index-1);
  F1_id=j+&N_F1*(index-1);
  F2_id=k+&N_F2*(index-1);
 end;
```

```
 end;
run;

Data F1_ran;
 do F1_id=1 to &N_B*&N_F1;
 F1_ran=SQRT(&V_F1)*rannor(0);
 output;
 end;
run;

Data F2_ran;
  do F2_id= 1 to &N_B*&N_F2 ;
   F2_ran=SQRT(&V_F2)*rannor(0);
  output;
  end;
run;

Data A_ran;
 do A_id=1 to &N_B*&N_A;
  A_ran=0.707*rannor(0);
  output;
 end;
run;

/*IF X IS AT LEVEL 1, USE THIS*/
Data A_ran;
 Do A_id=1 to &N_B*&N_A;
  x=ranbin(0, 1, 0.5);
  output;
 end;
run;

/*IF X IS AT LEVEL 2, USE THIS*/
Data F1_ran;
 do F1_id=1 to &N_F1;
  x=ranbin(0, 1, 0.5);
 output;
 end;
run;


/*IF X IS AT LEVEL 3, USE THIS*/
Data B_ran;
 do B_id = 1 to &N_B;
  x=ranbin(0, 1, 0.5);
 output;
 end;
run;


Proc sort data=B_F1_F2_A_total;
by F1_id;
run;
```

```
Data B_F1ran_F2_A;
merge B_F1_F2_A_total F1_ran;
by F1_id;
run;

Proc sort data=B_F1ran_F2_A;
by F2_id;
run;

Data B_F1ran_F2ran_A;
merge B_F1ran_F2_A F2_ran;
by F2_id;
run;

Proc sort data= B_F1ran_F2ran_A;
by A_id;
run;

Data all;
merge B_F1ran_F2ran_A A_ran;
y=0.1+0.5*x+B_ran+F1_ran+F2_ran+A_ran;
run;

proc mixed data=all;
C1ass B_id F1_id F2_id;
model y= x /solution;
random intercept /subject= B_id;
random intercept /subject= F1_id;
random intercept /subject= F2_id ;
ODS OUTPUT SolutionF=F_F1_F2 ;
ods output covparms=R_F1_F2;
run;

proc mixed data=all covtest;
class B_id F1_id;
model y= x/solution;
random intercept /subject= B_id;
random intercept /subject= F1_id;
Ods output solutionF=F_F1;
Ods output covparms=R_F1;
run;

Data F_F12_int (rename=(estimate=int_F12 stderr=std_i_F12) );
set F_F1_F2 ;
if effect='Intercept';
drop effect df probt tvalue;
run;

Data F_F12_x (rename=(estimate=x_F12 stderr=std_x_F12) );
set F_F1_F2 ;
if effect='x';
drop effect df probt tvalue;
run;
```

```
Data F_F1_int (rename=(estimate=int_F1 stderr=std_i_F1));
set F_F1;
if effect='Intercept';
drop effect df probt tvalue;
run;

Data F_F1_x (rename=(estimate=x_F1 stderr=std_x_F1));
set F_F1;
if effect='x';
drop effect df probt tvalue;
run;

Data fixed_1;
merge f_F12_int f_F12_x f_F1_int f_F1_x;
run;

proc append out=Fixed force;
run;

Data R_B_F12 (rename=(estimate=VB_F12 stderr=std_VB_F12));
set R_F1_F2;
if covparm="Intercept" & subject="B_id";
drop covparm subject zvalue probz;
run;

Data R_F1_F12 (rename=(estimate=VF1_F12 stderr=std_VF1_F12));
set R_F1_F2;
if covparm="Intercept" & subject="F1_id";
drop covparm subject zvalue probz;
run;

Data R_F2_F12 (rename=(estimate=VF2_F12 stderr=std_VF2_F12));
set R_F1_F2;
if covparm="Intercept" & subject="F2_id";
drop covparm subject zvalue probz;
run;

Data R_A_F12 (rename=(estimate=VA_F12 stderr=std_VA_F12));
set R_F1_F2;
if covparm="Residual";
drop covparm subject zvalue probz;
run;

Data R_B_F1 (rename=(estimate=VB_F1 stderr=std_VB_F1));
set R_F1;
if covparm="Intercept" & subject="B_id";
drop covparm subject zvalue probz;
run;

Data R_F1_F1 (rename=(estimate=VF1_F1 stderr=std_VF1_F1));
set R_F1;
if covparm="Intercept" & subject="F1_id";
drop covparm subject zvalue probz;
run;
```

```
Data R_A_F1 (rename=(estimate=VA_F1 stderr=std_VA_F1));
set R_F1;
if covparm="Residual";
drop covparm subject zvalue probz;
run;

Data random_1;
merge R_B_F12 R_F1_F12 R_F2_F12 R_A_F12
      R_B_F1 R_F1_F1 R_A_F1;
run;

Proc append out=Random force;
run;
```

APPENDIX C

SAS PROGRAM FOR GENERATING LONGITUDINAL CROSS-CLASSIFIED DATA

```
/*&PSI IS THE VARIANCE AND COVARIANCE OF STUDENT RANDOM EFFECT; &T IS
THE VARIANCE OF SCHOOL RANDOM EFFECTS; &NSCH IS THE NUMBER OF SCHOOLS;
&NSTU IS THE NUMBER OF STUDENTS PER SCHOOL; &MR IS THE MOBILITY RATE */

/*Generate Student random effects U0j and U1j*/
DATA A (TYPE=CORR);
input _Name_ $ _type_ $ U0 U1 ;
datalines;
UO CORR 1 .354
U1 CORR .354 1
;
run;

PROC FACTOR N=2 OUTSTAT=FACOUT;
RUN;

DATA PATTERN; SET FACOUT;
IF _TYPE_='PATTERN';
DROP _TYPE_ _NAME_;
RUN;

PROC IML;
 USE PATTERN;
 READ ALL VAR _NUM_ INTO F;
 F=F`;
 IF &PSI=1 THEN  Var={.1 0, 0 .05};
  ELSE IF &PSI=2 THEN VAR={.2 0, 0 .1};
 U0U1=RANNOR(J(&NSCH*&NSTU, 2, 0));
 U0U1=U0U1`;
 U0U1=F*U0U1;
 STD=SQRT(VAR);
 U0U1=(STD * U0U1)`;
 CREATE U0U1 FROM U0U1 [COLNAME={U0 U1}];
 APPEND FROM U0U1;
RUN;
QUIT;

/*Create school effect and Z*/
Data sch_effect;
 do school=1 to &NSCH;
  Z=rannor(0);
  V=SQRT(&T)*rannor(0);
 output;
 end;
run;
```

```
/*Create student file*/
Data stu_effect;
 do stu_id=1 to &N ;
  set U0U1;
  W=RANNOR(0);
 output;
 end;
run;

/*Create student membership at T1*/
Data student1;
 do schccmT1=1 to &NSCH;
     do student=1 to &NSTU;
        schhlmt1=schccmt1;
       output;
       end;
 end;
run;

Data stu_id_all;
 do stu_id=1 to &N;
 output;
 end;
run;

Data student1;
 merge student1 stu_id_all;
run;

/*Create student membership at T2*/
Data student2;
  set student1;
  do until (schccmt2 ne schccmt1);
     if student<=int(&NSTU*&MR) then schccmT2=int(&NSCH*ranuni(0)+1);
         schhlmt2=schhlmt1;
  output;
  end;
run;

Data student2;
 set student2;
  where schccmT2 ne schccmt1;
  if student > int(&NSTU*&MR) then schccmt2=schccmt1;
 run;

/*Create student membership from T1 to T4*/
Data student1_4;
 set student2;
 schccmt3=schccmt2;
 schccmt4=schccmt2;
 schhlmt3=schhlmt1;
 schhlmt4=schhlmt1;
 x1=0;
 x2=1;
```

```
 x3=2;
 x4=3;
 run;

 /*Transpose data*/
 Data student1_4long;
  set student1_4;
  array xvar [4] x1-x4;
  array schccm [4] schccmt1-schccmt4;
  array schhlm [4] schhlmt1-schhlmt4;
  do i=1 to 4;
   x=xvar[i];
   school_c=schccm[i];
   school_h=schhlm[i];
   output;
end;
drop i x1-x4 schccmt1-schccmt4 schhlmt1-schhlmt4;
run;

/*Merge student membership from T1 to T4 and student effect*/
Data student ;
 merge student1_4long stu_effect;
 by stu_id;
run;

/*Merget student effect and school effect*/
proc sql;
  create table sch_stu as
  select  *
  from student, sch_effect
  where student.school_c=sch_effect.school;
quit;

/*Create outcome*/
data all;
 set sch_stu;
 y=.1+.5*W+.5*Z+.5*X+.5*W*X+.5*Z*X+U0+U1*X+V+.63*rannor(0);
 drop school;
run;
```

APPENDIX D

R SYNTAX OF CCREMS FOR LONGITUDINAL CROSS-CLASSIFIED DATA

```
ccm = lmer(y~x+w+z+w*x+z*x+(x|stu_id)+(1|school_c),XX)
hlm = lmer(y~x+w+z+w*x+z*x+(x|stu_id)+(1|school_h),XX)
ccmres = summary(ccm)
hlmres = summary(hlm)
```

VITA

Name: Wen Luo

Address: Department of Educational Psychology

c/o Dr. Victor Willson

Texas A&M University, M.S. 4225

College Station, TX 77843-4225

Email Address: rowenaluo@gmail.com

Education: B.S., English, Shanghai Jiao Tong University, 2000

M.A., Linguistics, Shanghai Jiao Tong University, 2003

Ph.D., Educational Psychology, TAMU, 2007