December 2017

The Impact of Inappropriate Modeling of Cross-Classified Data Structures on Random-Slope Models

Feifei Ye
University of Pittsburgh, feifeiye.s@gmail.com

Laura Daniel
Lhd613@gmail.com

Recommended Citation

The Impact of Inappropriate Modeling of Cross-Classified Data Structures on Random-Slope Models

Feifei Ye
University of Pittsburgh
Pittsburgh, PA

Laura Daniel
Independent Consultant
Pittsburgh, PA

Previous studies that explored the impact of misspecification of cross-classified data structure as strictly hierarchical are limited to random intercept models. This study examined the effects of misspecification of a two-level, cross-classified, random effect model (CCREM) where both the level-1 intercept and slope were allowed to vary randomly. Results suggest that ignoring one of the crossed factors produced considerably underestimated standard errors for: 1) the regression coefficients of the level-1 predictor; 2) the inappropriately modeled predictor associated with the misspecified crossed factor; and 3) and their interaction. This misspecification also resulted in a significant inflation of the level-1 residual variances and the intercept and slope variance components across the levels of the remaining crossed factor in hierarchical linear model.

Keywords: Hierarchical linear model, cross-classified random effect model, Monte Carlo study

Introduction

Multilevel datasets in educational and social studies may have cross-classified, not purely nested, higher level sampling units, which prevents the use of hierarchical linear model. Examples of such data structure include students cross-classified by school and neighborhood (Raudenbush & Bryk, 2002), such as students in the same school coming from different neighborhoods and students in the same neighborhood going to different schools. Cross-classification also occurs in a longitudinal study when students have different math teachers at different grade levels, such as student math achievement cross-classified by the student and the teacher at a certain grade level.

Feifei Ye is an Assistant Professor of Research Methodology in the Department of Psychology in Education. Email her at: feifeiye.s@gmail.com.

The cross-classified random effects model (CCREM) is used to properly account for the cross-classified data structure. For example, the cross-classification of students in schools and neighborhoods can be modeled by a two-level CCREM, with both schools and neighborhoods considered level-2 units. In the pure hierarchical model in which neighborhoods are nested in schools, a three-level HLM is needed to separate the school (at level-3) and neighborhood (at level-2) effects, while assuming students in the same neighborhood all go to the same school.

With the recent development in computer software that can perform CCREM, a growing number of applied researchers have called upon the method to correctly model cross-classified data structures. A search of recent education research (1994-2014) using ERIC via EBSCO, using the keyword "cross-classified" indicated that 36 peer-reviewed empirical studies have used CCREM. However, given the prevalence of cross-classified data structures in educational research, only 36 studies in the last 20 years seems to be a small number.

In addition to infrequent use of the CCREM, there is also evidence that its use is, sometimes, consciously avoided by social scientists. In a study of neighborhood effects on educational achievement, Ainsworth (2002) removed subjects from the analysis that moved into new neighborhoods during the course of the study. Ma and Wilkins (2002) studied students' science achievement growth between the 7th and 12th grades using HLM to control for middle school clustering but ignoring high school clustering. More recently, Witherspoon and Ennett (2011) examined rural youths' developmental trajectories of self-reported grades, affective outcomes, and behavioral educational outcomes from 6th to 12th grade. The authors did not consider the nesting of adolescents within schools for two reasons, the reported complication of students changing schools over the course of the study and the small number of 9 schools. However, they did not use the school membership as a predictor in the model to account for the clustering of students in a school.

There are several suspected reasons that researchers may be hesitant to use CCREMs. First, there may be difficultly in ensuring the model prerequisites have been satisfied. In order for the use of a CCREM to be justified, both of the level-2 units for which a subject is cross-classified have to be randomly selected from larger populations. In the case where each of the cross-classified units are not randomly selected from larger populations, there has been some debate on whether the use of CCREM is appropriate. For example, Teitler and Weiss (2000) used a CCREM to examine the sexual behavior of youths who are cross-classified by census tract and school. Unless both census tract and school were randomly selected, the findings cannot be considered generalizable across all census tracts and schools in the U.S. Second, accurate information may not be readily available regarding

cluster membership in cross-classified collective structures. Third, the existence of CCREM may not be well known, or avoided due to its complexity.

The hesitance to use CCREMs may lead to misspecification of the cross-classified data. It was shown using empirical or simulated datasets misspecifying CCREM may result in misleading conclusions (Fielding, 2002; Goldstein, 1994; Luo & Kwok, 2009; Meyers & Beretvas, 2006). There are two approaches to misspecifying the cross-classified data structure, HLM-delete and HLM-complete (Meyers & Beretvas, 2006). For example, for students who were nested within a cross-classification of middle and high schools, HLM-delete omitted subjects who did not attend the main middle school that fed into their particular high school. This resulted in a strictly hierarchical dataset with students nested in middle schools which were nested in high schools, restricting the generalizability of the findings only to subjects who are not cross-classified (e.g., those who attended the main middle school). HLM-complete utilized all subjects but ignored the middle school clustering. Thus, students were only nested in high schools and the middle schools were not modeled as a separate level.

Meyers and Beretvas (2006) compared the two HLM approaches and CCREM when modelling test scores from students who were cross-classified by middle and high schools using the 1988 National Educational Longitudinal Study (NELS). They found that the fixed parameter estimates and their standard errors, the level-1 residual variance and its standard error were all similar between the HLM and CCREM models. However, the estimated between high school variance differed substantially between the models. The HLM-Delete model had the highest value followed by the HLM-Complete model, with the CCREM presenting the lowest between high school variance. This suggests in the HLM models the between middle school variance was masked to produce an inflated between high school variance.

Meyers and Beretvas (2006) replicated their real data analysis of the 1988 NELS data with a simulation study. Five factors were included in their design: correlation between the residuals of two cross-classified factors, number of feeder middle schools, number of levels of cross-classified units, average middle school size, and intraclass correlation (ICC) values. The CCREM model had students (level-1) nested within a cross-classification of middle and high schools (level-2), while the HLM model had students (level-1) nested within high schools (level-2), ignoring the middle school clustering. Both the CCREM and HLM models included three predictors, a student variable, a middle school variable, and a high school variable. However, while the CCREM modeled both the middle and high school characteristics as level-2 predictors, the HLM purposely modeled the middle school

characteristic which is a level-2 variable on level-1, as a student level characteristic. Results suggested that although the fixed parameter estimates were not affected as a result of misspecification, the standard errors for the middle school predictor that was included on the first level were. They found the relative biases of the standard error estimates under the CCREM model were all acceptable. However, most of the relative biases of the standard error values in the HLM model were intolerably high and negative. Furthermore, they found the between high school variance parameter was overestimated when the model incorrectly ignored the middle school clustering.

Luo and Kwok (2009) extended the above study by examining the impact of misspecification of CCREMs in a three-level model with two random factors crossed at the top level and at the intermediate level respectively. They found that ignoring one of the crossed factors biased the variance component estimates and standard errors of the fixed effects regression coefficients. The variance components of adjacent levels were overestimated and the variance component of the remaining crossed factor was underestimated. Further, misspecification resulted in underestimation of the standard error of the regression coefficient associated with a predictor of the ignored crossed factor and overestimation of the standard error of the regression coefficient of a predictor at a lower level.

Shi, Leite, and Algina (2010) assessed the effect of omitting the random interaction effect in CCREMs on parameter estimate and standard errors. No bias was found for the fixed effects. For random effects, variances at level-2 were affected but not those at level-1.

These three simulation studies examining the misspecification of CCREM focused on the random intercept model where the slope of the student (level-1) predictor remains constant at the cross-classified levels (e.g., middle school and high school). It is common for the effect of student predictors to vary across schools, and researchers are more interested in whether the effect is predicted by the school level predictors (i.e., cross-level interaction). Consequently a comparison of CCREM and HLM models with random slopes and intercepts warrants further investigation.

Studies in which the impact of misspecification of cross-classified datasets was explored can be compared to those exploring the impact of omitting a level in a purely hierarchical model (see, e.g., Moerbeek, 2004), which, similarly, were focused on random intercept models. This may be because analytical results are derivable with a closed-form solution for random intercept models, but not for random slope models (Van Landeghem, De Fraine, & Van Damme, 2005). When analytical results cannot be obtained, simulation studies are needed to evaluate the consequence of model misspecification. Random slope models are of interest in

cross-sectional research (e.g., Swanson & Stevenson, 2002) and longitudinal research. A random slope model is prevalent especially for the growth curve models as the changing slope across time is considered to vary across subjects. In summary, it is important to evaluate how ignoring cross-classified data structure affects the performance of fixed effects and variance components estimators in random slope models, and this study fills the gap. It is hypothesized ignoring cross-classified data structure will influence the standard error of fixed effects related to the misspecified level, and variance components of both random intercepts and slopes.

Methodology

Simulation Design

The design of this simulation mirrored the study of Meyers and Beretvas (2006) with one major difference: the slope of the level-1 predictor was modeled as randomly varying across both the middle and high schools. Four factors were manipulated: correlation between the level-2 residuals (0, 0.40); the intraclass correlation (ICCs) (0.05, 0.15, 0.25); the number of cross-classified units, (i.e., the number of middle schools and high schools, 30 and 50); and the average middle school size (20, 40). This resulted in a total of 24 conditions; for each condition, 2,000 datasets were generated.

Most of the simulation conditions in Meyers and Beretvas (2006) were adopted, except the added conditions with ICC = 0.25 as large ICC is common in with national longitudinal databases in education (Hedges & Hedberg, 2007). This study may be the first in which the impact of ignoring one level in multilevel analysis in random slope models is examined. It was considered important to keep the simulation conditions similar to the previous literature (Meyers & Beretvas, 2006; Luo & Kwok, 2009; Shi et al., 2010) so the results could be directly compared. As found when generating values, the distribution of predictors and the coefficients impacted the bias results. The simulation studies on CCREM, including Luo and Kwok (2009) and Shi et al. (2010) were based on similar parameter values as Meyers and Beretvas.

Unlike Meyers and Beretvas (2006), the number of feeder middle schools into high schools was held constant at two. Refer to their study for a detailed explanation of the way in which middle school students were organized into varying high schools. Meyer and Beretvas found the number of feeder middle schools did not affect outcome biases measures. Shi et al. (2010) came to a similar conclusion in their study of omitting random interaction effect in the cross-classified random

effects models. Different numbers of feeder schools resulted in different degrees of sparseness in the generated cross-classified cells. For example, with 30 high schools and 30 middle schools, there would be a total of 900 cells, and 840 (93.33%) cells in the two-feeder condition and 810 (90%) cells in the three-feeder condition would be empty. The degree of sparseness was found not to influence the bias outcome.

The correlation between the level-2 random effects for the middle and the high schools was set to be either 0 or 0.4. The zero correlation condition was chosen as a baseline model to compare to the 0.4. Meyers and Beretvas (2006) chose 0.4 to mimic the fairly related data structures found in many applied datasets (e.g. students who attend low socioeconomic middle schools are likely to attend low socioeconomic high schools as well).

The number of middle schools and high schools were either both 30 or both 50. The number of students in each middle school was randomly generated, either from a normal distribution with a mean of 20 and a standard deviation of 2 or from a normal distribution with a mean of 40 and a standard deviation of 2.

The three levels of ICCs were 0.05, 0.15, and 0.25. Meyers and Beretvas (2006) chose 0.05 and 0.15 to represent small and moderate ICCs as their examination of applied studies and textbook examples suggested that conditional ICCs ranged from 0.009 to 0.24, ($M = 0.066$, $SD = 0.0682$). The 0.25 was added here to represent large ICCs. Hedges and Hedberg (2007) examined achievement scores in longitudinal surveys with national probability samples and found that the average ICC was about 0.22 for all schools across Grades K-12, higher than the widely-used guidelines of 0.05-0.15.

Model Used to Generate Data

Data were using a two-level cross-classified model where students were cross-classified by middle school and high school and both level-1 intercepts and slopes were set to be random. Following the notation of Raudenbush and Bryk (2002), the level-1 equation of the CCREM model is:

$$Y_{i(jk)} = \pi_{0(jk)} + \pi_{1(jk)} X_{i(jk)} + e_{i(jk)}, \quad e_{i(jk)} \sim N\left(0, \sigma^2\right) \tag{1}$$

where $Y_{i(jk)}$ represents a student's achievement score, $\pi_{0(jk)}$ is the adjusted mean of students who had zero values on the student level predictor, $X_{i(jk)}$, and attended the same middle and high school combination. $\pi_{1(jk)}$ is the regression coefficient of the student level predictor. The level-1 residual, $e_{i(jk)}$, is the difference with which the

student's score varies from the cluster's adjusted mean, and has a unity variance. The level-2 equations are:

$$\begin{cases} \pi_{0(jk)} = \gamma_{000} + \gamma_{010}Z_j + \gamma_{020}W_k + b_{0j0} + c_{00k} \\ \pi_{1(jk)} = \gamma_{100} + \gamma_{110}Z_j + \gamma_{120}W_k + b_{1j0} + c_{10k} \end{cases} \tag{2}$$

The intercept model is the same as in Meyers and Beretvas (2006). The level-1 intercept is a function of an intercept, γ_{000}, a high school predictor, W_k, a middle school predictor, Z_j, plus the residual that is decomposed into two separate components for middle and high schools, b_{00j} and c_{00k}, respectively. However, unlike Meyers and Beretvas, the slope, $\pi_{1(jk)}$, of the level-1 predictor, $X_{i(jk)}$, is now modeled as randomly varying across middle and high schools, reflected in the b_{1j0} and c_{10k} terms. The random slope was also predicted by W_k and Z_j, resulting with two cross-level interaction terms.

The intercept (γ_{000}) was set to be at 100 and the coefficients of the predictors in the model ($\gamma_{100}, \gamma_{010}, \gamma_{020}, \gamma_{120}, \gamma_{110}$) were set to be 0.5 to replicate the study by Meyers and Beretvas (2006). The predictors (W_k, Z_j, X_{jk}) were all generated from normal distributions ($M = 0$, $SD = 10$). The mean of these distributions were set to be zero because in models with interaction terms, predictors should be centered to avoid multicollinearity. The lower-order effects (X, Z, W) were interpretable as the linear slope of a predictor given the other predictors at zero values. In applied research, researchers often need to make a decision regarding centering the student level predictor, X, as grand mean or group mean centered. Here, setting the mean of X to be zero in each combination of a middle school and a high school will produce similar results while centering with grand or group mean, thereby reducing the possible confounding effects introduced by the different centering approaches.

The CCREM model incorporated five variance components of b_{0j0}, b_{1j0}, c_{00k}, c_{10k}, and $e_{i(jk)}$ The residuals, b_{0j0} and b_{1j0} represented the amount of variation among middle schools in their adjusted mean (intercepts) and among the slopes of the student level predictor, after accounting for the middle and high school predictors. Similarly, c_{00k} and c_{10k} were the amount of variation among the high schools in the adjusted mean (intercepts) and the slopes of the student level predictor, after accounting for the middle and high school predictors. The variances of b_{00j}, c_{00k}, and their covariance depend on the values of ICC and correlation.

Table 1. Covariance matrix of random effects in data generation model for different correlation and ICC values

Correlation	ICC	Generating distribution of random intercepts and slopes
0.00	0.05	$\begin{pmatrix} b_{0j0} \\ b_{1j0} \\ c_{00k} \\ c_{10k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.0556 & 0 & 0 & 0 \\ 0 & 0.0278 & 0 & 0 \\ 0 & 0 & 0.0556 & 0 \\ 0 & 0 & 0 & 0.0278 \end{pmatrix} \right]$
	0.15	$\begin{pmatrix} b_{0j0} \\ b_{1j0} \\ c_{00k} \\ c_{10k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.2143 & 0 & 0 & 0 \\ 0 & 0.1072 & 0 & 0 \\ 0 & 0 & 0.2143 & 0 \\ 0 & 0 & 0 & 0.1072 \end{pmatrix} \right]$
	0.25	$\begin{pmatrix} b_{0j0} \\ b_{1j0} \\ c_{00k} \\ c_{10k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5000 & 0 & 0 & 0 \\ 0 & 0.2500 & 0 & 0 \\ 0 & 0 & 0.5000 & 0 \\ 0 & 0 & 0 & 0.2500 \end{pmatrix} \right]$
0.40	0.05	$\begin{pmatrix} b_{0j0} \\ b_{1j0} \\ c_{00k} \\ c_{10k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.0556 & 0 & 0.0222 & 0 \\ 0 & 0.0278 & 0 & 0 \\ 0.0222 & 0 & 0.0556 & 0 \\ 0 & 0 & 0 & 0.0278 \end{pmatrix} \right]$
	0.15	$\begin{pmatrix} b_{0j0} \\ b_{1j0} \\ c_{00k} \\ c_{10k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.2143 & 0 & 0.0857 & 0 \\ 0 & 0.1072 & 0 & 0 \\ 0.0857 & 0 & 0.2143 & 0 \\ 0 & 0 & 0 & 0.1072 \end{pmatrix} \right]$
	0.25	$\begin{pmatrix} b_{0j0} \\ b_{1j0} \\ c_{00k} \\ c_{10k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5000 & 0 & 0.2000 & 0 \\ 0 & 0.2500 & 0 & 0 \\ 0.2000 & 0 & 0.5000 & 0 \\ 0 & 0 & 0 & 0.2500 \end{pmatrix} \right]$

Presented in Table 1 are the generating distribution of the residuals in random intercepts and slopes (b_{0j0}, b_{1j0}, c_{00k}, c_{10k}) by correlation and ICC values. Using τ as the symbol for variance, the ICC is calculated as

$$\frac{\tau_{b_{0j0}}}{\tau_{b_{0j0}} + \tau_{c_{00k}} + \sigma^2}$$

for students attending the same middle school and

$$\frac{\tau_{c_{00k}}}{\tau_{b_{0j0}} + \tau_{c_{00k}} + \sigma^2}$$

for students attending the same high school. With $\tau_{b_{0j0}} = \tau_{c_{00k}}$ and $\sigma^2 = 1$, solve the values of the variances and then covariances (calculated as the square root of the product of corresponding variances and the correlation) for different conditions. Thus, for conditions with zero correlation, the variances of b_{0j0} and c_{00k} were set at 0.0556 for ICC = 0.05, at 0.2143 for ICC = 0.15, and at 0.50 for ICC = 0.25. For conditions with 0.4 correlation, the variances of b_{0j0} and c_{00k} were set with the same values of the corresponding conditions with zero correlation, while the covariance between b_{0j0} and c_{00k} was set at = 0.0222 for ICC = 0.05, 0.0857 for ICC = 0.15, and 0.2 for ICC = 0.25. The variances of b_{1j0} and c_{10k} were set to be half of the variances of b_{0j0} and c_{00k} as the variances of slopes are often smaller than those of intercepts (Raudenbush & Liu, 2001). The covariances between random intercepts and random slopes (b_{0j0} and b_{1j0}; c_{00k} and c_{10k}) were set to zero. Even though random intercepts and random slopes may be correlated in applied research, it was not considered in order to compare results to previous studies by introducing only variance components of random slopes. Introducing intercept-slope covariance may confound such comparison.

Data Generation and Analysis

Cross-classified datasets were generated using equations (1) and (2). First, we generated 50 (or 30 depending on the number of middle schools) 1 × 4 vectors from multivariate normal distributions as in Table 1. Combinations of middle schools and high schools were generated following the condition with two middle school feeders in Meyers and Beretvas (2006). The residuals of middle schools were then sorted into an ascending order. For all middle schools except the one with the largest residual, sixty percent of a middle school's students were sent to the high school paired with this middle school, and the rest were sent to the high school paired with the middle school with the next higher residual. For the middle school

466

with the largest residual, 40% of its students were sent to the high school paired with the middle school with the smallest residual.

Then, 2000 data sets were generated for each of the 24 simulation conditions in SAS 9.1.3. For each of the generated data set, SAS PROC MIXED was used to perform the CCREM analyses and HLM analyses using the full maximum likelihood estimation. The HLM addressed the high school clustering, but ignored the middle school clustering. The HLM included the middle school predictor and its interaction with the student predictor as the student level predictors. The level-1 equation of HLM model is

$$Y_{ij} = \pi_{0j} + \pi_{1j} X_{ij} + \pi_{2j} Z_{ij} + \pi_{3j} X_{ij} Z_{ij} + e_{ij} \tag{3}$$

This first level shows how the student's score, Y_{ij}, is a function of the intercept, π_{0j}, plus a student predictor, X_{ij}, and its weight, π_{1j}, as well as a middle school predictor erroneously included on the first level, Z_{ij}, and its weight, π_{2j}, the interaction between the two first level predictors, $X_{ij}Z_{ij}$, weighted by π_{3j}, plus an error term, e_{ij}, which captures the deviation from the student's score from the high school mean conditioned on the three predictors. The level-2 equations for the HLM are

$$\begin{cases} \pi_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \\ \pi_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j} \\ \pi_{2j} = \gamma_{20} \\ \pi_{3j} = \gamma_{30} \end{cases} \tag{4}$$

The level-1 intercept, π_{0j}, is predicted by an overall mean for all students conditioned on all predictors, γ_{00}, a high school predictor, W_j, its weight, γ_{01}, plus the error term, u_{0j}, which is the difference from a high school's adjusted mean to the overall mean for all high schools. The slopes, π_{1j}, are predicted by an overall slope, γ_{10}, plus a high school predictor, W_j, and its weight, γ_{11}, plus the error term, u_{1j}. The slope for the middle school predictor, π_{2j}, was held constant, as was the slope of the interaction term between the student and the middle school predictor.

The HLM analyses incorporated variance components of three residuals: u_{0j}, u_{1j}, and e_{ij}. The first residual, u_{0j} represents the amount of variation in the intercepts, or the mean achievement across high schools that remained unexplained after accounting for the predictors at the first and second levels. u_{1j} stands for the amount of residual variation in the slopes of the student level predictor across high schools,

and e_{ij} is the amount of student level variation that exists after controlling for the predictors. For comparison with CCREM estimates, only the variance components of intercepts and slopes between high schools and level-1 residuals were investigated. HLM analysis did not contain the middle school variance component, and thus could not be compared with CCREM on these estimates.

Outcome Measures

The outcome measures included relative biases of parameter estimates and standard error estimates of fixed effects and variance components. The relative bias of parameter estimates, $B(\hat{\theta})$, was calculated using equation (5), where $\bar{\hat{\theta}}_r$ is the mean of the rth parameter estimate across the 2,000 replication and θ_r is the actual value of the rth parameter (Hoogland & Boomsma, 1998).

$$B\left(\bar{\hat{\theta}}\right) = \frac{\bar{\hat{\theta}}_r - \theta_r}{\theta_r} \tag{5}$$

The relative bias of the standard errors (SEs), $B\left(\bar{\hat{S}}_{\hat{\theta}_r}\right)$ was calculated using equation (6), where $\bar{S}_{\hat{\theta}_r}$ is the mean standard error across the 2,000 replications and $S_{\hat{\theta}_r}$ is the standard deviation of the parameter estimates (i.e., empirical standard error) (Hoogland & Boomsma, 1998).

$$B\left(\hat{S}_{\hat{\theta}_r}\right) = \frac{\bar{\hat{S}}_{\hat{\theta}_r} - S_{\hat{\theta}_r}}{\hat{S}_{\hat{\theta}_r}} \tag{6}$$

Analysis

Descriptive statistics were used to examine whether the bias was acceptable for each condition; that is, whether the relative parameter bias was less than 0.05 and the relative standard error bias was less than 0.1 in absolute value (Hoogland & Boomsma, 1998). When biases were not acceptable in all simulation conditions for a parameter, analyses of variance (ANOVA) were conducted to determine which factor(s) affected the relative biases regarding this parameter. The outcome variables included the relative deviations of parameter estimates, calculated as the differences between the parameter estimates and the population parameters divided

by the population parameters, and the relative deviations of standard error estimates, calculated as the deviations of the sample standard error estimate from the empirical standard error divided by the empirical standard error.

The between-subjects factors in ANOVAs included the correlation between the level-2 residuals, zero and non-zero; the intraclass correlation, 0.05, 0.15, and 0.25; the number of cross-classified units, 30 and 50; and the average middle school size, 20 and 40. The within-subjects factor included the model type (CCREM and HLM). Due to the large number of replications, statistical significance was not examined. Instead, partial eta squared effect sizes, η_p^2, were computed as a measure of practical significance. Only effects that were practically significant with $\eta_p^2 > 0.01$ were interpreted, which was considered a very conservative cut-off value for an appreciable effect in similar simulation studies (Shi et al., 2010).
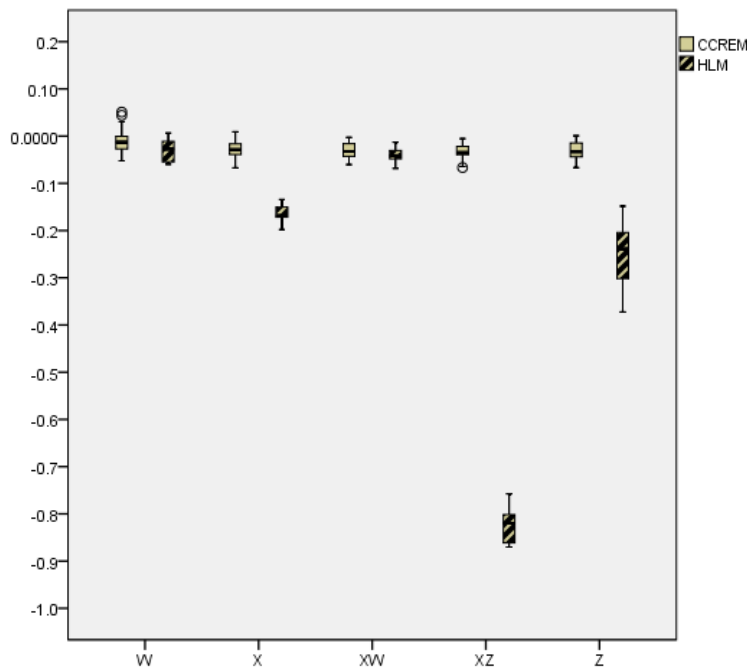
Figure 1. Relative standard error bias of coefficients of *W* (the high school predictor), *X* (the student level predictor), *Z* (the middle school predictor), and interactions between *X* and *W* and between *X* and *Z*

Results

Relative Parameter and Standard Error Bias of Fixed Effects

All relative parameter biases for fixed effect estimates met the criterion of Hoogland and Boomsma (1998) (< 0.05) for acceptable parameter bias amounts. All relative standard error biases for fixed effect estimates, as shown in Figure 1, met Hoogland and Boosma's criterion (< 0.10) for acceptable parameter bias amounts except the HLM estimates of coefficients of *X*, *Z*, and *XZ*.

Coefficient of X, the Student Level Predictor   For coefficients of *X* (γ_{100} for CCREM and γ_{10} for HLM), the CCREM standard error estimates were in the acceptable range for all conditions with biases ranging from -0.07 to 0.01 ($M = -0.03$, $SD = 0.02$). The HLM standard error estimates were unacceptable for all conditions with biases ranging from -0.20 to -0.13 ($M = -0.16$, $SD = 0.02$). Given the substantial bias found in the HLM standard error estimates, a 2 (correlation) × 3 (ICC) × 2 (number of cross-classified units) × 2 (average middle school size) between-subjects ANOVA was conducted on the relative deviations of HLM standard error estimates. Only the main effect of number of middle schools was significant $\left(\eta_p^2 = 0.012 \right)$. The other effects had $\eta_p^2 < 0.004$. Biases in conditions with 30 schools ($M = -0.18$, $SD = 0.15$) were larger in magnitude than those with 50 schools ($M = -0.15$, $SD = 0.09$).

Coefficient of Z, the Middle School Predictor   For the coefficient of *Z* (γ_{010} for CCREM and γ_{20} for HLM), CCREM standard error estimates were acceptable. The relative standard error bias ranged from -0.07 to 0.00 ($M = -0.03$, $SD = 0.02$). However, HLM standard error estimates were negatively biased with a range from -0.37 to -0.15 ($M = -0.25$, $SD = 0.06$). The between-subjects ANOVA conducted on the relative deviations of HLM standard error estimates indicated that the significant effects included the main effects of ICC $\left(\eta_p^2 = 0.08 \right)$ and middle school size $\left(\eta_p^2 = 0.14 \right)$. The other effects had $\eta_p^2 < 0.003$. Pairwise comparison suggested that biases for ICC = 0.25 ($M = -0.29$, $SD = 0.14$) were larger in magnitude than those for ICC = 0.15 ($M = -0.26$, $SD = 0.13$), which in turn were larger than those for ICC = 0.05 ($M = -0.20$, $SD = 0.13$). Biases were larger in magnitude when middle school size was 40 ($M = -0.30$, $SD = 0.12$) than 20 ($M = -0.20$, $SD = 0.14$).

Coefficient of XZ, the Cross-Level Interaction between the Student Level Predictor and the Middle School Predictor For the coefficient of *XZ* (γ_{110} for CCREM and γ_{30} for HLM), the CCREM standard error estimates were in the acceptable range for all conditions with the biases ranging from -0.07 to -0.01 (*M* = -0.03, *SD* = 0.02). The standard error biases of HLM estimates exceeded the acceptable level for all of the conditions ranging from -0.87 to -0.76 (*M* = -0.82, *SD* = 0.04). A between-subjects ANOVA on the relative deviations of HLM standard error estimates presented similar results as those for the coefficient of *Z* with the main effects of ICC $\left(\eta_p^2 = 0.24\right)$ and middle school size $\left(\eta_p^2 = 0.45\right)$ significant. The other effects had $\eta_p^2 < 0.008$. Pairwise comparison suggested that biases increased in magnitude for larger ICCs (*M* = -0.84, *SD* = 0.04 for ICC = 0.25; *M* = -0.83, *SD* = 0.04 for ICC = 0.15; *M* = -0.80, *SD* = 0.05 for ICC = 0.05). Biases were larger in magnitude when middle school size was 40 (*M* = -0.85, *SD* = 0.03) than 20 (*M* = -0.79, *SD* = 0.04).

In summary, the HLM model resulted with deflated standard errors of regression coefficients related to the student level predictor, the middle school predictor, and their interaction. This is not surprising, because the HLM model ignores the middle school level and disaggregates the middle school predictor at the student level.

Table 2. Relative bias of variance estimates when correlation between residuals was 0

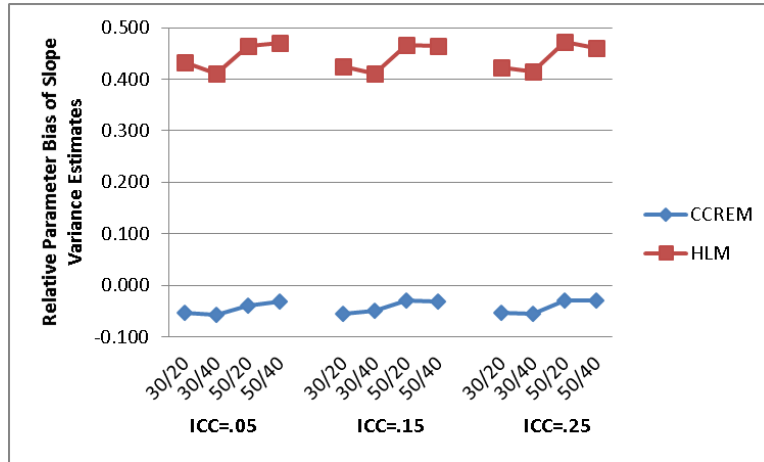| ICC	No. of schools	Middle school size	Variance of intercept across high schools		Variance of slope across high schools		Student-level residual variance	
			CCREM	HLM	CCREM	HLM	CCREM	HLM
0.05	30	20	-0.108	0.232	-0.054	0.432	0.000	1.219
0.05	30	40	-0.103	0.299	-0.057	0.411	0.002	1.260
0.05	50	20	-0.079	0.328	-0.039	0.464	0.001	1.255
0.05	50	40	-0.035	0.412	-0.031	0.471	0.000	1.283
0.15	30	20	-0.088	0.252	-0.055	0.424	0.000	4.711
0.15	30	40	-0.059	0.357	-0.050	0.411	0.000	4.836
0.15	50	20	-0.057	0.326	-0.029	0.466	-0.001	4.825
0.15	50	40	-0.045	0.398	-0.032	0.464	0.000	4.979
0.25	30	20	-0.066	0.267	-0.053	0.423	0.000	11.012
0.25	30	40	-0.055	0.353	-0.056	0.414	0.000	11.384
0.25	50	20	-0.038	0.364	-0.030	0.473	0.000	11.319
0.25	50	40	-0.042	0.399	-0.029	0.460	0.000	11.578
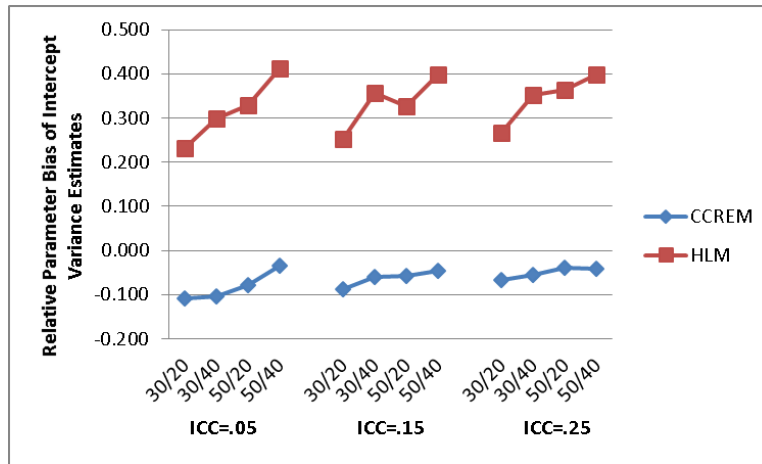
Figure 2. Relative parameter bias of variance estimates of intercepts across high schools for conditions with correlation = 0; the two numbers on the horizontal axis were number of schools / middle school size

Bias of the Random Effects Variance Components for Zero Correlation Conditions

As Meyers and Beretvas (2006) pointed out, when correlation between middle school and high school intercepts residual was not zero, the true value of the variance components was not known because the analysis using CCREM assumed that the correlation of cross-classified units was zero. Thus, only relative parameter bias is reported for the student level residual variance and variance of intercepts and slopes between high schools for conditions when correlation was zero (Table 2). For conditions when correlation was 0.4, compare the variance estimates of CCREM and HLM.

Variance of Level-1 Intercepts across High Schools The relative bias ranged from -0.108 to -0.035 for CCREM estimates and from 0.232 to 0.412 for HLM estimates across the 24 simulation conditions (Figure 2). The CCREM estimates in the conditions with larger sample sizes (50 schools and an average of 40 students per school) were acceptable with relative biases less than 0.05. CCREM estimates in the other conditions were slightly negatively biased. However, HLM estimates were positively biased in all conditions. A mixed effect ANOVA was conducted on the relative deviations of the parameter estimates. Only the main effect of model was significant $\left(\eta_p^2 = 0.45\right)$ with CCREM estimates having smaller

deviations (M = -0.06, SD = 0.37) than HLM estimates (M = 0.33, SD = 0.68). The other effects had $\eta_p^2 < 0.003$.

Figure 3. Relative parameter bias of variance estimates of slopes across high schools for conditions with correlation = 0; the two numbers on the horizontal axis were number of schools / middle school size
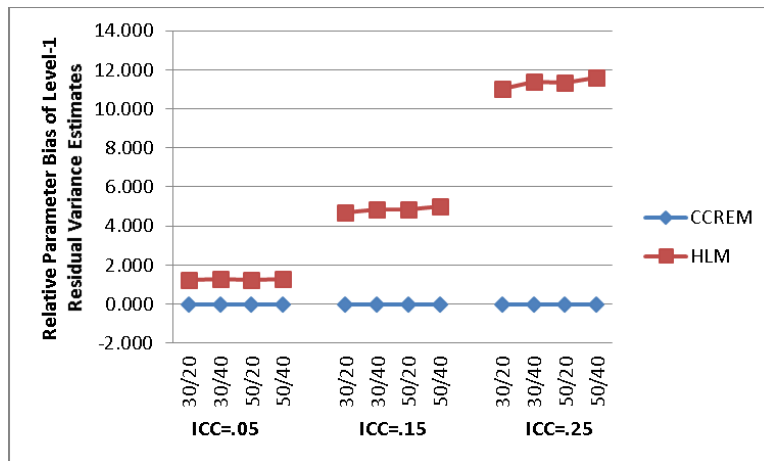
Figure 4. Relative parameter bias of level-1 residual variance estimates for conditions with correlation = 0; the two numbers on the horizontal axis were number of schools / middle school size

473

Variance of Level-1 Slope across High Schools   CCREM estimates of the slope variance across high schools were acceptable with the relative biases ranging from -0.057 to -0.029 (Figure 3). However, HLM estimates were unacceptable with biases ranging from 0.411 to 0.473. A mixed effect ANOVA on the relative deviations of the parameter estimates resulted with only one appreciable effect, the main effect of model $\left(\eta_p^2 = 0.76 \right)$. The other effects had $\eta_p^2 < 0.004$.

Student-Level Residual Variance   There was no bias in estimating student level residual variance using CCREM because the biases ranged from -0.0038 to -0.0007. However, the HLM model overestimated the student level residual variance with the bias ranging from 1.219 to 11.578. The mixed ANOVA indicated that the appreciable effects included: interaction effect between model and ICC $\left(\eta_p^2 = 0.79 \right)$, and main effects of model $\left(\eta_p^2 = 0.88 \right)$ and ICC $\left(\eta_p^2 = 0.79 \right)$. As shown in Figure 4, the degree of overestimation of HLM estimates increased with the value of ICC.

In summary, the HLM model inflated the variance components of the level-1 residual, the random intercept and slope across high schools. In other words, while omitting the middle school level, variance across middle schools was imposed at the student and high school level.

Table 3. Mean of random effect variance estimates when correlation between residuals was 0.4

ICC	No. of schools	Middle school size	Var of intercept across high schools		Var of slope across high schools		Student-level residual variance	
			CCREM	HLM	CCREM	HLM	CCREM	HLM
0.05	30	20	0.060	0.090	0.026	0.040	0.996	2.221
0.05	30	40	0.061	0.097	0.026	0.040	0.999	2.254
0.05	50	20	0.064	0.099	0.027	0.041	0.998	2.245
0.05	50	40	0.063	0.102	0.027	0.041	0.998	2.289
0.15	30	20	0.232	0.357	0.101	0.153	0.996	5.750
0.15	30	40	0.225	0.381	0.102	0.153	0.998	5.848
0.15	50	20	0.239	0.382	0.104	0.158	0.997	5.858
0.15	50	40	0.232	0.396	0.104	0.157	0.998	5.950
0.25	30	20	0.524	0.857	0.237	0.357	0.998	11.969
0.25	30	40	0.508	0.891	0.238	0.356	0.999	12.297
0.25	50	20	0.543	0.908	0.242	0.367	0.997	12.269
0.25	50	40	0.526	0.928	0.243	0.367	0.999	12.565
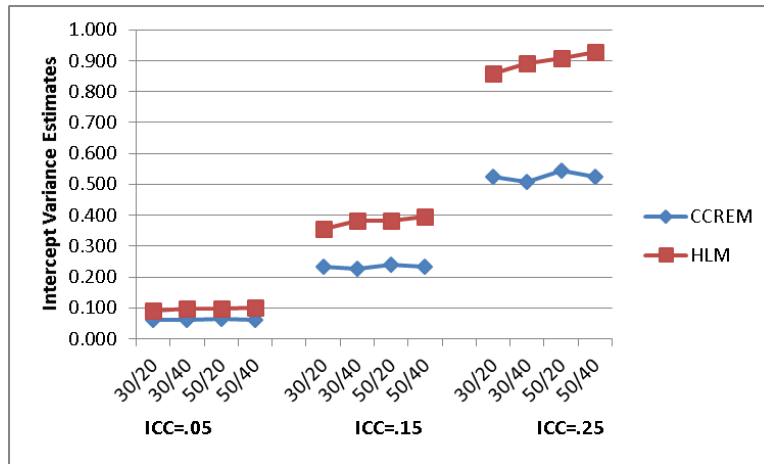
Figure 5. Variance estimates of intercepts across high schools for conditions with correlation = 0.4; the two numbers on the horizontal axis were number of schools / middle school size
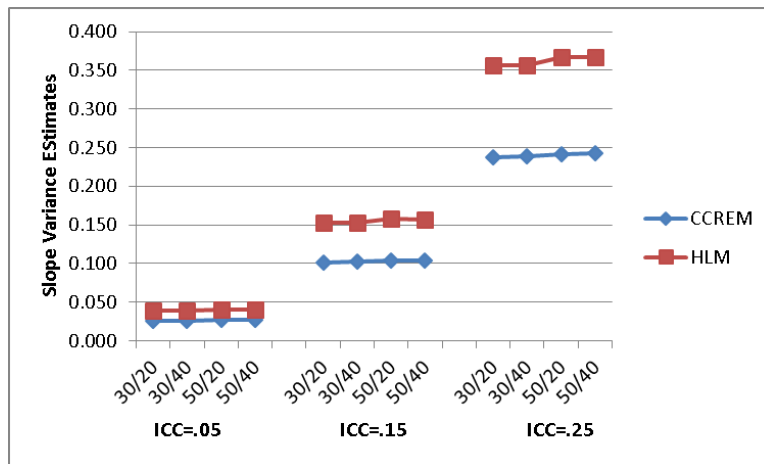
Figure 6. Variance estimates of slopes across high schools for conditions with correlation = 0.4; the two numbers on the horizontal axis were number of schools / middle school size

## Random Effects Variance Components for 0.4 Correlation Conditions

Compared in Table 3 are CCREM and HLM estimates of the variances of intercepts and slopes across high schools, and the student level residual variance for conditions with 0.4 correlation. The CCREM estimates were close to the generating

475

values of these variances. For variances of intercepts and slopes (Figures 5 and 6), the HLM estimates were larger than CCREM estimates, and the differences increased with larger ICC values. For student level residual variance (Figure 7), HLM estimates were almost twice as large as the CCREM estimates when ICC = 0.05, four times larger than the CCREM estimates when ICC = 0.15, and twelve times larger when ICC = 0.25. Note the variances of intercepts and slopes over the middle schools, the ignored crossed factor in HLM, were the same as those of the high schools in the data generation. It is evident that the variance contributed by the ignored crossed factor (i.e., the middle school) was imposed on the variance components of the other remaining crossed factor (i.e., the high school) and the student level when using HLM.

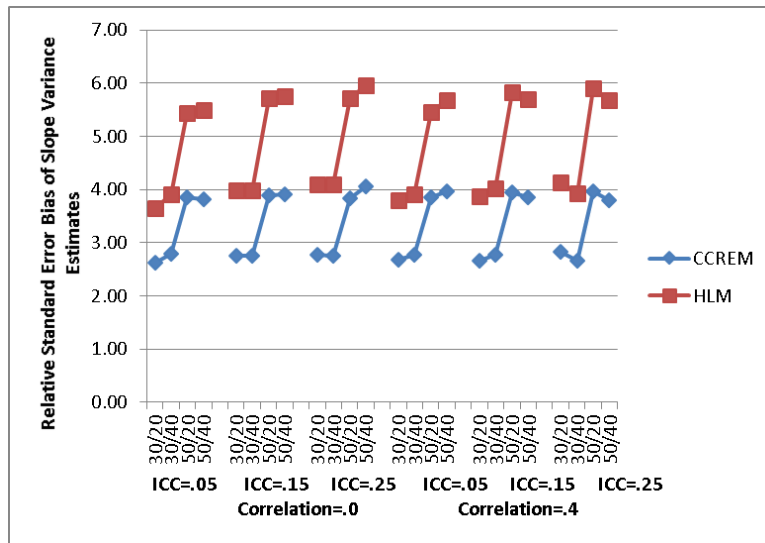## Relative Standard Error Bias of Random Effects Variance Components

Presented in Table 4 and Figures 8 and 9 are the relative standard error biases across the 24 conditions of estimates of intercepts and slopes variances. None of the CCREM and HLM standard error estimates were acceptable. For intercepts variances, CCREM standard error estimates had biases ranging from 2.44 to 3.91 ($M = 3.10$, $SD = 0.54$) and HLM standard error estimates had biases ranging from 3.48 to 7.60 ($M = 5.33$, $SD = 1.21$). For slope variances, CCREM standard error estimates had biases ranging from 2.62 to 4.06 ($M = 3.10$, $SD = 0.54$) and HLM standard error estimates had biases ranging from 3.64 to 5.96 ($M = 4.82$, $SD = 0.90$).
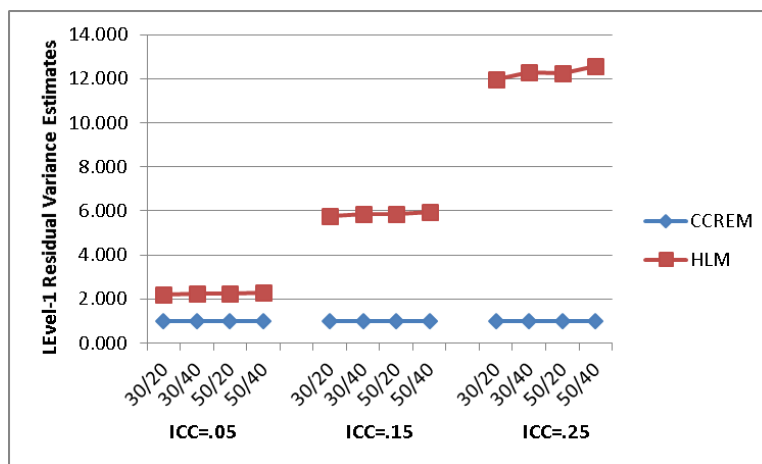
Figure 7. Level-1 residual variance estimates for conditions with correlation = 0.4; the two numbers on the horizontal axis were number of schools / middle school size
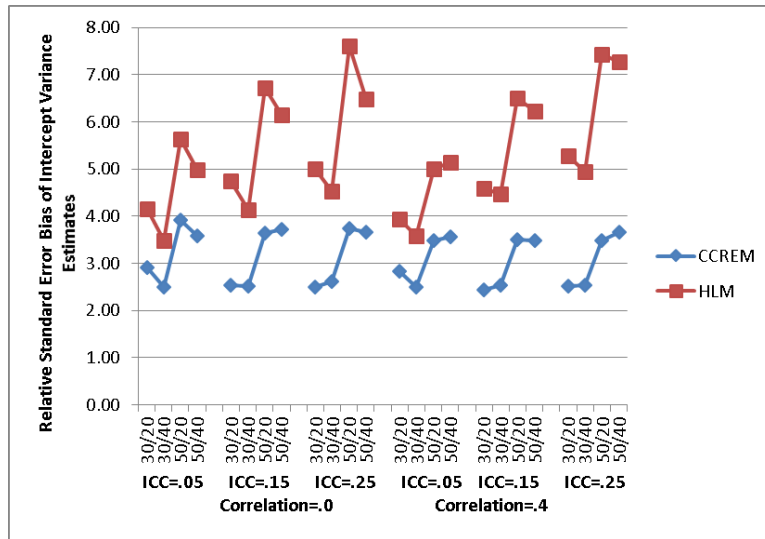
Figure 8. Relative standard error bias of variance estimates of intercepts across high schools; the two numbers on the horizontal axis were number of schools / middle school size

Figure 9. Relative standard error bias of variance estimates of slopes across high schools; the two numbers on the horizontal axis were number of schools / middle school size

Table 4. Relative standard error bias of random effect variance estimates

Correlation	ICC	No. of schools	Middle school size	SE of intercept variance across high schools		SE of slope variance across high schools	
				CCREM	HLM	CCREM	HLM
0.0	0.05	30	20	2.92	4.15	2.62	3.64
0.0	0.05	30	40	2.49	3.48	2.79	3.90
0.0	0.05	50	20	3.91	5.64	3.85	5.44
0.0	0.05	50	40	3.58	4.99	3.81	5.50
0.0	0.15	30	20	2.54	4.74	2.75	3.99
0.0	0.15	30	40	2.52	4.14	2.76	3.99
0.0	0.15	50	20	3.64	6.73	3.89	5.71
0.0	0.15	50	40	3.72	6.14	3.91	5.76
0.0	0.25	30	20	2.49	5.00	2.78	4.10
0.0	0.25	30	40	2.62	4.54	2.76	4.09
0.0	0.25	50	20	3.74	7.60	3.83	5.71
0.0	0.25	50	40	3.67	6.48	4.06	5.96
0.4	0.05	30	20	2.83	3.94	2.68	3.80
0.4	0.05	30	40	2.49	3.58	2.77	3.90
0.4	0.05	50	20	3.48	5.00	3.85	5.45
0.4	0.05	50	40	3.56	5.15	3.96	5.67
0.4	0.15	30	20	2.44	4.58	2.66	3.87
0.4	0.15	30	40	2.54	4.48	2.77	4.02
0.4	0.15	50	20	3.50	6.51	3.95	5.83
0.4	0.15	50	40	3.49	6.23	3.86	5.70
0.4	0.25	30	20	2.51	5.28	2.82	4.14
0.4	0.25	30	40	2.53	4.95	2.66	3.93
0.4	0.25	50	20	3.49	7.43	3.96	5.90
0.4	0.25	50	40	3.66	7.28	3.79	5.68

Mixed effect ANOVA was conducted on the relative deviations of standard error estimates for intercepts variances and slope variances. For intercepts variances, the appreciable effects included the interaction effects between model and ICC $\left(\eta_p^2 = 0.14\right)$, model and the number of schools $\left(\eta_p^2 = 0.06\right)$, model and middle school size $\left(\eta_p^2 = 0.01\right)$, and main effects of model $\left(\eta_p^2 = 0.63\right)$, ICC $\left(\eta_p^2 = 0.05\right)$, the number of schools $\left(\eta_p^2 = 0.23\right)$, and middle school size $\left(\eta_p^2 = 0.01\right)$. In Figure 9, we could see that HLM estimates were more positively biased than CCREM estimates. The differences in the bias between HLM and CCREM estimates were

larger for larger ICC values, 50 schools (vs. 30), and middle school size of 20 (vs. 40).

For slope variances, the effects that met the criterion for an appreciable effect included: the interaction effects between model and the number of schools $\left(\eta_p^2 = 0.06\right)$, and main effects of model $\left(\eta_p^2 = 0.64\right)$, and the number of schools $\left(\eta_p^2 = 0.29\right)$. Figure 9 shows that HLM estimates were more positively biased than CCREM estimates and such difference was larger for conditions with 50 schools (vs. 30), and middle school size of 20 (vs. 40).

Power of Detecting Nonzero Variance Components

Power was all 100% for hypothesis tests regarding slope variances across high schools for both CCREM and HLM estimates, but not for intercepts variances. Figure 10 presents power levels for hypothesis tests regarding intercepts variances across high schools. When ICC was 0.05, there was no difference in power between CCREM and HLM. For higher ICC values, the CCREM estimates had near one power level while HLM estimates had significantly lower power in conditions with middle school size of 20. The difference in the power was the largest for the conditions with the smallest sample sizes (30 schools with 20 students on average per school).
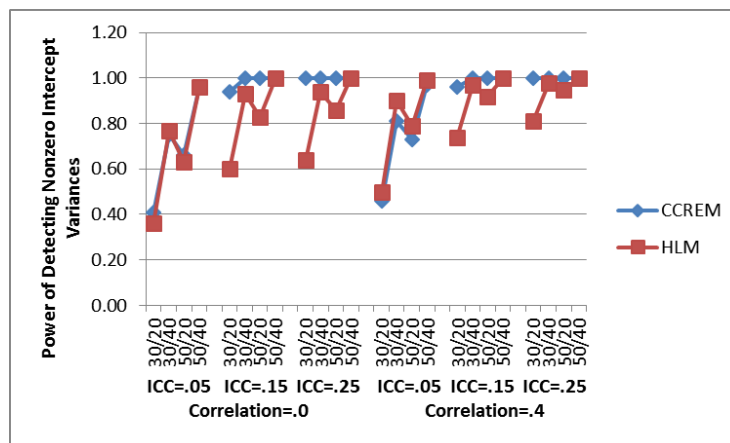
Figure 10. Power of detecting nonzero intercept variances using CCREM and HLM models; the two numbers on the horizontal axis were number of schools / middle school size

Conclusion

In this study, previous research of ignoring a crossed random factor was extended by examining its effect on CCREM models with both random intercepts and slopes. One kind of model misspecification of CCREM models was considered: misspecifying a cross-classified level-2 unit (e.g., middle school) and imposed its characteristics (e.g., the SES of the middle school) as a predictor at level-1 (e.g., student level). Our results were consistent to previous Monte Carlo research (Meyers & Beretvas, 2006; Luo & Kwok, 2009) in that employing HLM on cross-classified data resulted in 1) unbiased estimates of regression coefficients of all fixed effects; 2) unbiased standard errors of regression coefficient of the high school predictor (the remaining crossed factor); 3) underestimated standard errors of regression coefficient for the middle school predictor (the predictor misspecified as level-1 predictor); 4) overestimated level-1 residual variance; and 5) overestimated variance of random intercepts across high schools.

In the previous studies, a random intercept model was adopted in which the effect of the student level predictor was fixed across different middle and high schools. In this study, this limitation was freed, and the effect of the student level predictor was modeled as random across middle and high schools. In addition, two cross-level interaction effects were modeled in which the effect of the student level predictor was affected by a middle school predictor and a high school predictor. The addition of random slope changed the impact of model misspecification on the fixed effect regression coefficient of the student level predictor found in previous studies. Meyers and Beretvas (2006) found misspecified HLM produced unbiased standard error of the regression coefficient of the student level predictor. However, with the addition of the random slope to model in the current study, it was found this standard error was underestimated and magnitude of underestimation was larger when there were fewer cross-classified units (middle and high schools).

The focus in previous studies on random intercept models omitted the cross-level interaction between the student level predictor and the high school predictor, the interaction between the student level predictor and the middle school predictor, as well as the variance of random slopes of the student level predictor across high schools. In the current study, these omissions were taken into account and it was found model misspecification did not impact the regression coefficient or the standard error of the cross-level interaction between the student level predictor and the high school predictor. The reason may be that clustering of students in high schools was considered in the HLM. However, it was found for the interaction between the student level predictor and the middle school predictor, the standard

error was underestimated using HLM and the magnitude of underestimation was higher for larger ICC values and larger size schools. The dependence of underestimation on the size of the schools was also found by Meyers and Beretvas (2006). This underestimation may be due to the contribution of the per-cluster sample size to the design effect in cluster sampling. According to Snijders and Bosker (1999), the design effect is larger when the sample size per cluster is larger, and thus more bias is introduced when the homogeneity within a cluster is ignored. In addition, the design effect is larger when ICC is larger, thus explaining more biases for larger ICC values.

The results from this study suggest, across all conditions considered, the use of CCREM for cross-classified datasets produced more accurate results when compared to incorrectly using a hierarchical model. It appears that when a predictor is modeled on a level lower than it should be and a level of nesting is ignored, the parameter of the true predictor on that lower level, as well as its interaction with the erroneously modeled predictor, may have biased estimates. This result revealed the insidious effect that ignoring a level of nesting may have on the variables in the model that were actually correctly specified. The downward bias of the standard error for regression coefficients of predictors at level-1, predictors at the ignored level, and their cross-level interaction has important implications for applied researchers who are utilizing an HLM. This underestimation inflates the Type I error rate. Specifically, the degree of underestimation of HLM standard error estimates is the most pronounced for the cross-level interaction. The false identification of significant interaction effects could lead to excessive effort in exploring and explaining such interaction effects. When ICC was not substantial (e.g., at 0.05), the standard error estimates in the misspecified HLM models were still significantly underestimated. This suggests trying to model the cross-classification even when there is concern with the degree of dependence within a crossed factor.

The effect of misspecification on random intercepts variance found in previous studies applies to random slopes variance as well. In the current study, the variances of the level-1 intercepts and slopes across high schools were both overestimated in the misspecified HLM. However, this does not necessarily imply that misspecified HLM models would be more powerful in detecting the nonzero variance components. The standard error estimates of CCREM and HLM are both positively biased while HLM estimates have larger biases. With substantial clustering, CCREM is found to have more power in detecting the nonzero variance components, but HLM has similar power when the number of schools and school sizes are large. Caution is needed in using the standard error of an estimated

variance component to conduct a hypothesis test. The distribution of the estimated variance component converges only slowly to normality as the sample size increases, and thus a Wald z-test on variances is not recommended (Raudenbush & Bryk, 2002). With positively biased standard error estimates, CCREM still produces acceptable power when ICC suggests substantial clustering (ICC $\geq$ 0.15). Raudenbush and Bryk (2002) states that the software of HLM provided a better approach to hypothesis testing of variance components, and further studies may be needed in comparing their approach to Wald's z-test in hypothesis testing of variance components.

The biased variance estimates of HLM have serious consequences. If a level-2 crossed factor is ignored, and instead an HLM is implemented, the notion of how much student level variance remains unexplained after accounting for the predictors would be severely inaccurate. Using a CCREM greatly reduces this bias and provides an unbiased estimate of how much variation in the student level remains after controlling for the predictors. Moreover, the overestimation of variances of intercepts and slopes across the remaining crossed factor can lead to incorrect conclusions on the degree of remaining dependence after all predictors are controlled for. This may lead to the identification of a greater number of predictors than actually needed to explain the overestimated variances.

In summary, CCREM should be applied to cross-classified data to avoid inflated Type I error of some fixed effects and overestimated variances of level-1 intercepts and slopes. When CCREM cannot be applied to a cross-classified data due to lack of information of cluster membership, if the conditions are similar to those used in this study, ignoring a crossed factor will not impact the predictors of the correctly modeled crossed factor. Hypothesis tests regarding predictors at the lower level and associated with the ignored crossed factor will have inflated Type I error rates. Intercepts and slopes variances will be overestimated.

Limitations and Suggestions for Future Research

The generalizability of results from this study is restricted to the manipulated conditions of the four independent variables in this study and thus is not applicable to all research cases. Indeed, there were several limitations of this study that can be addressed by research methodologists in the future. The principle limitations of this study are contingent on the manipulation of the four factors. Furthermore, as Meyers and Beretvas (2006) purported, more research is needed when the number of cross-classified units (middle and high schools) is not identical. The covariance between random intercepts and random slopes could be nonzero. Last, only partial

cross-classification was considered in this study. As Luo and Kwok (2009) pointed out, cross-classification structure has important impact on the effect of model misspecification of cross-classified data. Different cross-classification structures should be evaluated in the future for random slope models.

References

Ainsworth, J. W. (2002). Why does it take a village? The mediation of neighborhood effects on educational achievement. *Social Forces, 81*(1), 117-152. doi: 10.1353/sof.2002.0038

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*(2), 225-246. doi: 10.1076/sesi.13.2.225.3435

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods & Research, 22*(3), 364-375. doi: 10.1177/0049124194022003005

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87. doi: 10.3102/0162373707299706

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26*(3), 329-367. doi: 10.1177/0049124198026003003

Luo, W., & Kwok, O.-m. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research, 44*(2), 182-212. doi: 10.1080/00273170902794214

Ma, X. & Wilkins, J. L. M. (2002). The development of science achievement in middle and high school: Individual differences and school effects. *Evaluation Review, 26*(4), 395-417. doi: 10.1177/0193841x02026004003

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41*(4), 473-497. doi: 10.1207/s15327906mbr4104_3

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39*(1), 129-149. doi: 10.1207/s15327906mbr3901_5

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

Raudenbush, S. W., & Liu, X.-F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6*(4), 387-401. doi: 10.1037/1082-989x.6.4.387

Shi, Y., Leite, W., & Algina, J. (2010). The impact of omitting the interaction between crossed factors in cross-classified random effects modeling. *British Journal of Mathematical and Statistical Psychology, 63*(1), 1-15. doi: 10.1348/000711008x398968

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

Swanson, C. B., & Stevenson, D. L. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessment. *Educational Evaluation and Policy Analysis, 24*(1), 1-27. doi: 10.3102/01623737024001001

Teitler, J. O., & Weiss, C. C. (2000). Effects of neighborhood and school environment on transitions to first sexual intercourse. *Sociology of Education, 73*(2), 112-132. doi: 10.2307/2673240

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*(4), 423-434. doi: 10.1207/s15327906mbr4004_2

Witherspoon, D., & Ennett, S. (2011). Stability and change in rural youths' educational outcomes through the middle and high school years. *Journal of Youth and Adolescence, 40*(9), 1077-1090. doi: 10.1007/s10964-010-9614-6