Asia Pacific Education Review
2008, Vol. 9, No.2, 206-220.

Closing the Gap: Modeling within-school Variance Heterogeneity in School Effect Studies

Junyeop Kim
Charles Drew University
USA

Kilchan Choi
University of California, Los Angeles
USA

Effective schools should be superior in both enhancing students' achievement levels and reducing the gap between high- and low-achieving students in the school. However, the focus has been placed mainly on schools' achievement levels in most school effect studies. In this article, we focused our attention upon the school-specific achievement dispersion as well as achievement level in determining effective schools. The achievement dispersion in a particular school can be captured by within-school variance in achievement (σ^2). Assuming heterogeneous within-school variance across schools in hierarchical modeling, it is possible to identify school factors related to high achievement levels and a small gap between high- and low-achieving students. By analyzing data from the TIMMS-R, we illustrated how to detect variance heterogeneity and how to find a systematic relationship between within-school variance and school practice. In terms of our results, we found that schools with a high achievement level tended to be more homogeneous in achievement dispersion, but even among schools with the same achievement level, schools varied in their achievement dispersion, depending on classroom practices.

Keywords: school effect, variance heterogeneity, achievement gap, hierarchical modeling, latent variable regression

One of the fundamental questions that most school effect studies have continuously addressed is whether schools make a difference in student achievement, and if so, how much of the student achievement can be attributable to schools' effort. Regarding this question, most researchers have agreed that schools do have a measurable impact on student achievement, even though the source and the magnitude of the school effect are still heavily debated (Rumberger & Palardy, 2003).

Using a simple Hierarchical Model (HM), one can successfully show how much of the total variation in achievement comes from the student level (within-school variance, σ^2) and how much comes from the school level (between-school variance, τ). Many studies have found that between-school variance is much smaller than within-school variance. For example, using High School and Beyond (HS&B) data, Lee and Bryk (1989) found that about 19% of the total variation in student math achievement was attributable to school differences.

More complicated HMs can be used to discover the source of these within- and between-school variances. Because school effect studies are usually focused on identifying effective schools after controlling for student

Junyeop Kim, Assistant professor at the Charles Drew University, USA; Kilchan Choi, National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles, USA.

Correspondence concerning this article should be addressed to Junyeop Kim, Charles Drew University, 1748 E, 118th Street. Building N, Los Angeles, CA 90059, United States. e-mail: junyeopkim@cdrewu.edu

background characteristics, or on finding out school practices that are effective in increasing student achievement, between-school variance (τ) plays an important role. Substantial τ is evidence of a school's contribution to student outcome, indicating the magnitude of variation among schools in their achievement levels (Raudenbush & Bryk, 2002). On the other hand, there has been little discussion on within-school variance (σ^2) in school effect studies.

We argue in this article that σ^2 can provide valuable information regarding effective schools because school effectiveness can be determined not only by student achievement levels, but also by the dispersion of student achievement within a school. Given that all schools try to increase their students' levels of achievement, it is clear that successful schools should have a smaller degree of variability in their student achievement levels. Additionally, these achievement levels themselves should be higher because a high average achievement level *plus* smaller within-school variation indicates that the school has successfully directed a majority of its students to a certain level. In other words, effective schools should be superior in both increasing students' achievement levels *and* reducing the gap between high- and low-achieving students in the school. The former can be captured in common HM and has been addressed in many school effect studies. The latter— the dispersal of student achievement within a school— can be captured through within-school variance by assuming that σ^2 varies across schools with careful examination of variance heterogeneity in HM.

The purpose of this study was to illustrate how to detect variance heterogeneity and find a systematic relationship between within-school variance and school practices. If certain school practices are related to smaller within-school variance, this could provide important information in the way in which school practice can have an equalizing effect on student performance.

Data

Data from the Third International Mathematics and Science Study-Repeat (TIMSS-R) was used for this study. TIMSS-R is an international study of math and science achievement conducted by the International Association for the Evaluation of Educational Achievement (IEA) in 1999 (Eugenio & Julie, 2001). The target population was eighth-grade students, and 38 countries participated in the study. The dataset contains student, teacher, and school background data, as well as student math and science achievement scores. More information can be found at the TIMSS website, *www.timss.org*.

Due to the fact that the purpose was not international comparison, we used a data from a single country (Republic of Korea) and a single content area score (math achievement). In TIMSS-R, this math achievement score was equated across countries using Item Response Theory and rescaled to have a mean of 500 and a standard deviation of 100. We used a subsample from the larger TIMSS-R sample. The final sample contains 5,583 students in 143 Korean schools. The average math achievement was 590.62, and the standard deviation was 77.60 - almost 1 *SD* above the international average achievement level with smaller variation.

Earlier studies, using the same dataset, reported significant student- and school-level variables affecting student achievement (Park, Park, & Kim, 2001; Yang & Kim, 2003). According to these studies, most of the variation in math achievement was attributable to difference among students (95.6% in Park, Park, & Kim and 93.2% in Yang & Kim) and between-school variation (τ) contributed only less than 7% of the total variation. Significant student level correlates of higher math achievement identified in these studies include higher family SES, positive attitudes toward math, after school time management (taking extra math lessons and spend less time watching TV or playing with friends). In addition, the average socioeconomic status (SES) level and school location were closely related to achievement at the school level. These significant variables are selected to specify achievement models in our study (see Tables 1 and 2). However, it is possible that some important covariates are omitted from the model specification, as is generally recognized in most observational studies. This problem is discussed later in relation to the variance heterogeneity.

For student background characteristics, student gender (GENDER), parents' highest education level (PED), and the home educational resources index (HOMERSC) were used. HOMERSC is a composite variable that the IEA created, based on students' responses regarding educational resources

Table 1
Descriptive Statistics for Student-level Variables

Name	Description	Category/Scale	Freq.(%)	Mean	*SD*
GENDER	Student gender	0:Boy / 1:Girl	2872 (51.4) 2711 (48.6)	0.49	0.50
PED	Parents highest education level	0: No primary grad. 1: Primary 2: Secondary/college 3: University	719 (12.9) 768 (13.8) 2677 (47.9) 1419 (25.4)	1.86	0.94
HOMRSC	Home educational resources index	0(Low) / 1(Medium) / 2(High)	283 (5.1) 4471 (80.1) 829 (14.8)	1.10	0.44
LESSON	Take extra math lesson outside school more than 1 hour/week	0 (No) / 1 (Yes)	3288 (58.9) 2295 (41.1)	0.41	0.49
AOFREQ	Teacher explains rules at the beginning of new topic	1 (never)~4 (always)	229 (4.1) 741 (13.3) 1602 (28.7) 3011 (53.9)	3.32	0.86
STDUSEBOD	How often student use board	1 (never)~4 (always)	932 (16.7) 2511 (45.0) 1252 (22.4) 888 (15.9)	2.38	0.94
MATATT	Positive attitude towards math	0(Low) / 1(Medium) / 2(High)	1487 (26.6) 3594 (64.4) 502 (9.0)	0.82	0.57
TIMEPLY	Spend 3 or more hours/day watching TV/video or playing with friends	0 (No) / 1 (Yes)	2083 (37.3) 3500 (62.7)	0.63	0.48

Table 2
Descriptive Statistics for School-level Variables

Name	Description	Category/Scale	Freq.(%)	Mean	*SD*
URBAN	Urban schools	0 (No) / 1 (Yes)	69 (48.3) 74 (51.7)	0.52	0.50
SUBURB	Suburban schools	0 (No) / 1 (Yes)	89 (62.2) 54 (37.8)	0.38	0.49
MPED	School mean PED	Continuous		1.85	0.32
USEBOD*	How often teacher use board	1 (never)~4 (always)		3.05	0.17

Note. USEBOD is entered for variance modeling.

at home, such as computers, the student's own desk, dictionary and number of books. In terms of student's experience outside school, extra math lessons outside school (LESSON) and time spent on non-academic activities such as watching TV or playing with friends (TIMEPLY) were selected. Student responses to the frequency of teacher's advance organizer use (AOFREQ) and how often students used the board (STDUSEBOD) were also selected to model the impact of the classroom experience. Finally, student's positive attitudes toward mathematics (MATATT) was selected to check the impact of student motivation.

Some student-level variables are aggregated to the school level to measure the contextual effects and school practice effects. The school mean of parents' education level (MPED) can be used to measure the contextual effect of SES, which will be discussed in the results section. School location (rural, suburban or urban) is entered as a dummy variable to estimate the school location effects.

Math teacher's use of the board (USEBOD) was selected to model within-school heterogeneity for illustrative purposes. We assume that this variable, when aggregated to the school level, describes an important classroom practice. If a teacher uses the board more frequently, students in that class will share the same instructional experience more often, and as a result, math achievement for those students will become more similar. Based on this assumption, USEBOD was entered to explain variance heterogeneity. If this variable has an equalizing effect, schools in which teachers use the board more frequently should have a smaller variation in student achievement. Findings regarding this variable may suggest an equalizing effect; however, such a result may require closer theory-based investigation before further generalization. We also used school mean achievement levels as a predictor for modeling variance heterogeneity, because we hypothesize that in effective schools, achievement should be both high on average and narrow in dispersion. In other words, we will illustrate that even among schools with a similar average achievement level, the gap between high- and low-achievers can vary across schools depending on school's instructional characteristics.

Methods and Models

A common practice in HM application is to assume that all errors at level-1 are drawn from an identical distribution, that is, $r_{ij} \sim N(0, \sigma^2)$. It is reported that the estimation of fixed effects and their standard errors does not change substantially when this assumption does not hold and σ^2 varies randomly (Kasim & Raudenbush, 1998). Because of the robustness of this assumption, school effect studies rarely pay attention to the possibility of heterogeneous variance. However, level-1 variance may differ across schools and can give valuable information regarding the equalizing effect of certain school practices.

However, one needs to specify the mean structure carefully before modeling the dispersion heterogeneity because variance heterogeneity can result from model misspecification. Bryk and Raudenbush (1988) pointed out that in randomized experiments, heterogeneous variance across groups can be viewed as an indicator that shows the possibility of treatment and aptitude interaction. Similarly in a multilevel situation, heterogeneity may be caused by model misspecification, either by omitting an important level-1 variable or by erroneously fixing a level-1 predictor slope (Raudenbush & Bryk, 2002). Another source of variance heterogeneity comes from differences in school characteristics and is of central interest in the effort to identify effective schools. Note that omission of the level-2 variables in the mean structure is less problematic. Kim and Seltzer (2006) pointed out that in multisite studies that use multilevel model, differences in school level characteristics may work as another source of level-1 variance heterogeneity but is not necessarily associated with the model misspecification problem. That is, the omission of school-level correlates of level-1 variance heterogeneity in the mean structure model does not affect the inference on variance heterogeneity. School effect studies that center on the school factors that reduce the achievement gap should attend to these correlates of dispersion. However, it should be pointed out that modeling heterogeneous variance using school level covariates does not compensate for the model misspecification problem and finding a systematic relationship between level-1 variance and school characteristics does not reduce the bias in fixed effects estimates caused by the misspecification of mean structure.

If we find heterogeneity in level-1 variances after establishing the final model, the next step is to model this residual variance to see whether there is a systematic pattern. Variance homogeneity can be tested by computing chi-

Table 3
Results from Unconditional Model

Fixed Effect	Estimate	*s.e.*	*t*-ratio	*p*-value
Grand mean	590.22	1.912	308.64	0.000
Variance Components	Estimate		χ^2	*p*-value
Between school variance	379.81		507.70	0.000
Within school variance	5652.50			
Homogeneity of level-1 var. test			177.63	0.02

square statistics for standardized dispersion (see Raudenbush & Bryk, 2002, pp. 263-265, for example). After checking the variance heterogeneity, the next step would be to examine the distribution of variances and set up a regression model to find a relationship with school characteristics. Our specific models and their development are discussed below.

First, we fit a fully unconditional model to decompose the total variance into student and school levels. The results showed that the grand mean math achievement was 590.22, between-school variance was 379.81, and within-school variance was 5652.50. These results indicate that only about 6.3% of the total variance is attributable to school differences and the remaining 93.7% of the total variance comes from individual differences among students within schools. The test of homogeneous variance rejected the homogeneous variance assumption. The results are summarized in Table 3.

As noted above, variance heterogeneity could occur by omission of an important level-1 variable or by fixing the level-1 slope that is in fact varying across schools. To make sure this was not the case, we fit a series of HM as described below.

First, all eight level-1 variables entered in the model and random variation was allowed only for intercept (Model 1, Random intercept ANCOVA model). The level-1 homogeneous variance test for this model still rejected the null hypothesis that level-1 residual errors are drawn from an identical distribution. Following Raudenbush and Bryk (2002), we checked the variability of level-1 slopes across schools and found that LESSON, AOFREQ, and TIMEPLY effects varied significantly across schools. Therefore, in Model 2, we allowed random variation for intercept and the three slopes. Furthermore, in this model, school location and the average education level of parents (MPED) were entered to model the intercept (adjusted grand mean). The chi-square test for this model also rejected the homogeneous level-1 variance assumption. In Model 3, school location and MPED were entered for the three random slopes specified in Model 2, as well as for the intercept. This was the final mean structure model. The homogeneous variance assumption was again rejected in this model. Therefore, we moved to the heterogeneous variance model, keeping the mean structure, as specified in Model 3. The results for Models 1 through 3 are summarized in Table 4. The statistics package HLM6 (Raudenbush, Bryk, Condon, & Cheong, 2004) was used to fit the three models described above.

In the heterogeneous variance model, level-1 variance is assumed to vary across schools. Therefore, we posed a school-specific within-school residual variance, σ^2_j for school j. The first step in our variance modeling was to check whether schools with higher achievement levels had smaller σ^2_j. Modeling within-school residual variance using school mean achievement has two important implications. First, in relation to our definition of an effective school, using average achievement level as a predictor for achievement gap enables us to answer questions regarding the overall tendency, that is, whether high achievement schools are in general effective schools under our definition – a significant negative effect of achievement level on level-1 variance will confirm this question. Next, to make a causal statement that schools' instructional settings either reduce or magnify achievement gap, we need a strong assumption that the entry achievement gap was equal across schools. Since the data are not from a randomized experiment, this equal entry gap assumption is both too strong and difficult to

Table 4

Result Summary for Model 1 to Model 3 with Homogeneity of Level-1 Variance Test

	Model 1			Model 2			Model 3		
Fixed Effects	Estimate	*s.e.*	*t* (*p*-value)	Estimate	*s.e.*	*t* (*p*-value)	Estimate	*s.e.*	*t* (*p*-value)
For adjusted grand mean, β_{0j}									
Adjusted grand mean, γ_{00}	590.46	1.49	396.24(0.00)	590.41	1.31	450.69(0.00)	590.94	1.32	447.67 (0.00)
Urban schools, γ_{01}				19.73	5.40	3.66 (0.00)	18.44	5.5	3.34 (0.00)
Suburban schools, γ_{02}				16.46	5.41	3.04 (0.00)	14.69	5.49	2.67 (0.00)
School average parent ed. Level, γ_{03}				11.84	4.21	2.81 (0.00)	14.92	4.46	3.33 (0.00)
Gender contrast, γ_{10}	-3.33	3.06	-1.08 (0.27)	-2.93	3.10	-0.95 (0.35)	-2.77	3.14	-0.88 (0.37)
Parent highest ed. Slope, γ_{20}	6.92	1.22	5.68 (0.00)	6.24	1.23	5.09 (0.00)	6.12	1.23	4.94 (0.00)
Home resource slope, γ_{30}	32.86	2.64	12.44 (0.00)	32.02	2.64	12.11 (0.00)	32.16	2.63	12.21 (0.00)
For extra outside lesson sloope, β_{4j}									
Adjusted mean effect, γ_{40}	21.73	2.30	9.44 (0.00)	20.62	2.34	8.81 (0.00)	21.22	2.34	9.06 (0.00)
Urban schools, γ_{41}							-10.56	9.66	-1.09 (0.27)
Suburban schools, γ_{42}							-4.68	9.78	-0.47 (0.63)
School average parent ed. Level, γ_{43}							-15.21	7.21	-2.10 (0.03)
For 'teacher explains rules at the beginning' slope, β_{5j}									
Adjusted mean effect, γ_{50}	14.94	1.33	11.22 (0.00)	14.75	1.32	11.17 (0.00)	14.53	1.3	11.17 (0.00)
Urban schools, γ_{51}							2.44	7.05	0.34 (0.72)
Suburban schools, γ_{52}							2.62	7.05	0.37 (0.71)
School average parent ed. Level, γ_{53}							-10.25	4.98	-2.05 (0.03)
Student use board' slope, γ_{60}	6.48	1.12	5.78 (0.00)	6.61	1.11	5.95 (0.00)	6.63	1.09	6.05 (0.00)
Positive attitude toward math slope, γ_{70}	27.38	1.68	16.24 (0.00)	27.73	1.69	16.40 (0.00)	27.74	1.69	16.41 (0.00)
For 'spend more than 3 hrs. playing/TV' slope, β_{8j}									
Adjusted mean effect, γ_{80}	-12.34	2.05	-5.99 (0.00)	-13.18	2.05	-6.42 (0.00)	-13.24	2.01	-6.57 (0.00)
Urban schools, γ_{81}							-13.92	7.49	-1.85 (0.06)
Suburban schools, γ_{82}							-14.27	7.49	-1.90 (0.05)
School average parent ed. Level, γ_{83}							4.29	6.53	0.65 (0.51)
Variance Components	Estimate	χ^2	*p*-value	Estimate	χ^2	*p*-value	Estimate	χ^2	*p*-value
Within-school (σ^2)	4417.46			4312.19			4313.77		
Between (intercept, τ_{00})	205.83	397.40	0.000	135.33	284.60	0.000	133.57	283.13	0.000
Between (extra lesson slope, τ_{44})				198.45	184.12	0.009	169.02	174.50	0.019
Between (teacher explain rules slope, τ_{55})				65.51	196.90	0.002	61.75	189.59	0.003
Between (3 or more hrs playing slope, τ_{88})				93.86	169.27	0.052	88.07	166.76	0.048
Homogeneity of level-1 var. test		174.37	0.033		175.30	0.030		183.37	0.010

justify. By controlling for achievement level, this assumption needs to be satisfied only among schools with similar achievement levels. To control for the achievement level, we used a latent variable regression technique (Raudenbush & Bryk, 2002; Seltzer, Choi, & Thum, 2003; Choi & Seltzer, in press), which essentially uses the unobserved latent variable (adjusted school mean, β_0 in this study) as a predictor for σ^2_j (Model 4). This latent variable modeling approach, instead of using observed mean achievement, enables us to avoid the attenuation problem of the regression coefficient, which is caused by measurement errors involved in observed variables.

An effective school, according to our definition, is a school with high achievement and small variation in achievement among its students. Therefore, to determine school effectiveness it is crucial to examine school characteristics and practices that can reduce student achievement variation even after controlling for school mean achievement. Our final model (Model 5) illustrates this point. Both β_0 (average achievement level) and USEBOD were entered to model σ^2_j. Therefore, a significant USEBOD effect will indicate that among schools with the same achievement level, schools in which teachers use the board more frequently have a smaller gap between high- and low-achieving students. Specifications of the final models are shown in equations (1), (2), and (3).

Note that at the student level, PED, HOMERSC and LESSON are grand mean centered and other level-1 variables are group mean centered. These grand mean centered variables are related to either SES or academic input from outside the school, and would be better controlled in school effect studies because variation in student achievement due to these variables cannot be attributable to school practice. This is a particularly reasonable approach if we can assume, for example, that a school's average achievement is high because most of its students take extra lessons outside school; then it would be more reasonable to adjust for the effect of these outside lessons when we evaluate the school's performance. By virtue of this level-1 centering, β_{0j} now represents the average math achievement of school *j*, controlling for parents' education, home educational resources, and extra math lessons. β_{1j} through β_{8j} capture the effect of corresponding variables, respectively—that is, the average increase/decrease of student achievement in school *j* when the value of the corresponding variable changes by one unit.

Achievement Model

Within school

$$Y_{ij} = \beta_{0j} + \beta_{1j}(GENDER_{ij}) + \beta_{2j}(PED_{ij}) + \beta_{3j}(HOMERSC_{ij}) + \beta_{4j}(LESSON_{ij}) + \beta_{5j}(AOFREQ_{ij}) + \beta_{6j}(USEBOD_{ij}) + \beta_{7j}(MATATT_{ij}) + \beta_{8j}(TIMEPLY_{ij}) + r_{ij},$$

$$r_{ij} \sim N(0, \sigma_j^2) \tag{1}$$

Between school

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(URBAN_j) + \gamma_{02}(SUBURBAN_j) + \gamma_{03}(MPED_j) + u_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30} \\
\beta_{4j} &= \gamma_{40} + \gamma_{41}(URBAN_j) + \gamma_{42}(SUBURBAN_j) + \gamma_{43}(MPED_j) + u_{4j} \\
\beta_{5j} &= \gamma_{50} + \gamma_{51}(URBAN_j) + \gamma_{52}(SUBURBAN_j) + \gamma_{53}(MPED_j) + u_{5j} \\
\beta_{6j} &= \gamma_{60} \\
\beta_{7j} &= \gamma_{70} \\
\beta_{8j} &= \gamma_{80} + \gamma_{81}(URBAN_j) + \gamma_{82}(SUBURBAN_j) + \gamma_{83}(MPED_j) + u_{0j},
\end{aligned}$$

$$\begin{pmatrix} u_{0j} \\ u_{4j} \\ u_{5j} \\ u_{8j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{04} & \tau_{05} & \tau_{08} \\ & \tau_{44} & \tau_{45} & \tau_{48} \\ & & \tau_{55} & \tau_{58} \\ & & & \tau_{88} \end{pmatrix} \right) \tag{2}$$

Dispersion Model

$$\sigma_j = d_0 + d_1(\beta_{0j} - \gamma_{00}) + d_2(USEBOD_j) + e_j, \qquad e_j \sim N(0, \delta^2) \tag{3}$$

At the school level, some βs are allowed to vary across schools, and school location and average PED level (MPED) are entered as predictors—also note that all the school- level variables are grand mean centered. By this grand mean centering, γ_{00} now captures the adjusted grand mean achievement level. γ_{01} and γ_{02} indicate how much urban and suburban schools did better/worse than the grand mean. γ_{03} requires special attention for interpretation—this fixed effect captures the contextual effect of parents' education levels. Since we already have adjusted for PED at the student level, γ_{03} captures, among students with similar parental education levels, how much extra advantage students receive in schools with a one-unit-higher mean PED level.

Because preliminary analysis found no variability in GENDER, PED, and HOMERSC effects across schools, these slopes are fixed at the school level. Therefore, γ_{10}, γ_{20}, and γ_{30} show the overall gender difference, PED effect, and HOMERSC effect, respectively. LESSON and AOFREQ slopes showed significant variability across schools, and these slopes are set to vary randomly across schools. γ_{40} captures the overall extra lessons effect. γ_{41} through γ_{43} show whether extra lessons are more effective in urban schools (γ_{41}), suburban schools (γ_{42}) or in schools with higher average SES levels (γ_{43}). γ_{50} through γ_{53} can be interpreted the same way as γ_{40} through γ_{43}. USEBOD and MATATT slopes are also fixed across schools. Therefore, γ_{60} and γ_{70} represent the overall USEBOD effect and MATATT effect, respectively. γ_{80} is the overall achievement difference between students spending 3 or more hours playing/watching TV and students spending less than 3 hours in those nonacademic activities. γ_{81} and γ_{82} show whether this difference is larger or smaller in urban schools (γ_{81}) and suburban schools (γ_{82}). Finally, γ_{83} shows whether the gap gets wider or narrower depending on school mean SES level.

As we specified in the level-1 model (equation 1), each school has its own within-school residual variance (σ^2_j), and σ^2_j now captures the dispersion of student achievement in school *j* after explaining the effect of student-level variables.

Before modeling the variance, we examined the distribution of σ^2_j[1] (left in Figure 1). The distribution seems positively skewed with one outlying school (school #142). Due to the fact that variance can only take positive values, it is a common practice to log-transform the estimated variance to fit the model (Raudenbush & Bryk, 2002). This transformation reduces the degree of the skew and enables the transformed value to take on a negative value. However, because log-transformation is a non-linear function, the interpretation becomes rather complex. Another option in this situation is a square root transformation. Even though this is also a non-linear transformation, the interpretation becomes straightforward, considering the fact that the square root of variance is the standard deviation. On the right hand side of Figure 1 the distribution after square root transformation is shown. Note that the transformed data are less skewed. The degree of the skew of the original scale is 1.11, whereas the degree of the skew of the transformed scale is substantially reduced to 0.52. Therefore, we fit the variance regression model using the square root of the variance as the outcome. Note also that, for a sensitivity check, we fit the same model without school #142, which seems to have outlying variance. The result was not substantially different. [2]

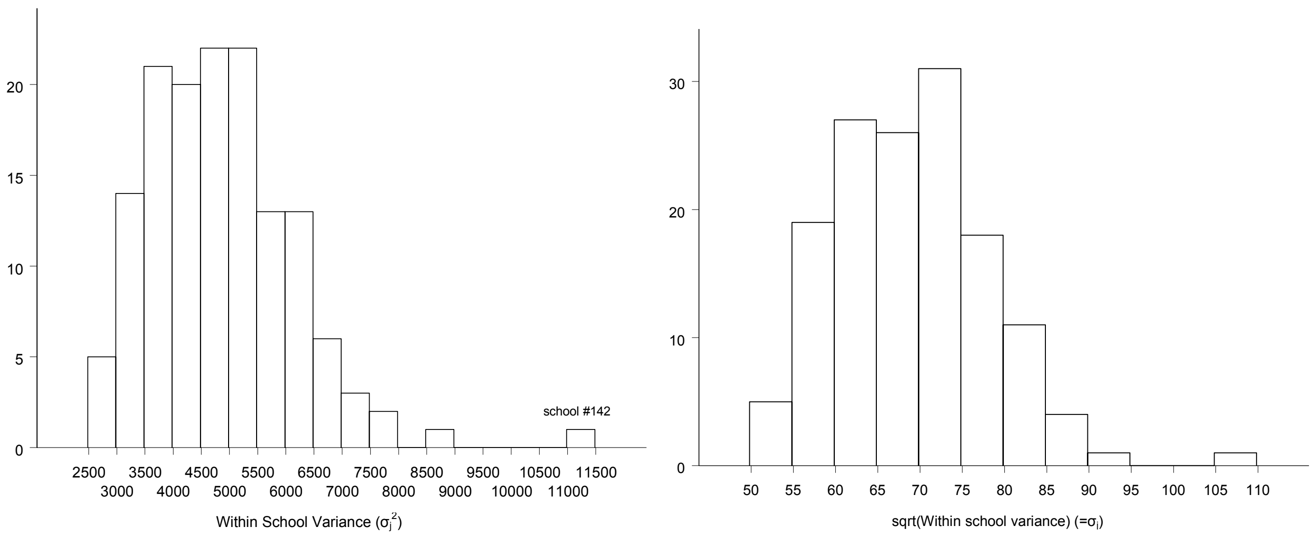

Figure 1. Comparing distributions of σ^2_j and σ_j.

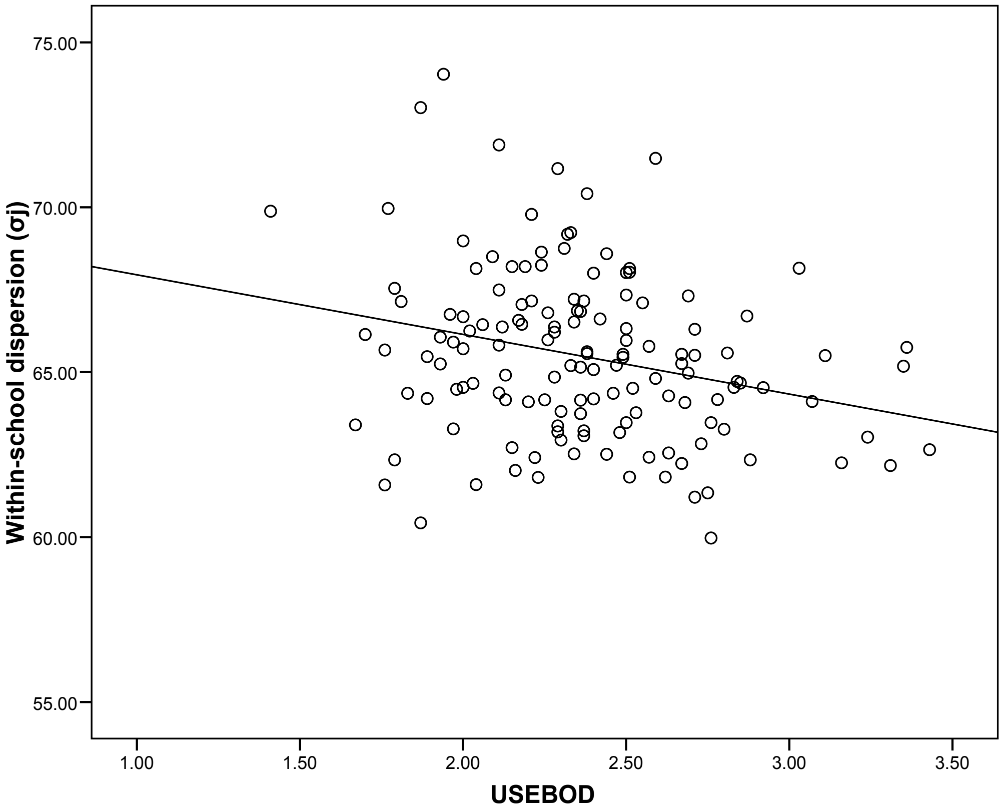

Figure 2. Relationship between σ_j and *USEBOD*

Furthermore, before placing a regression model on the dispersion, we checked the relationship between school-specific dispersion and the covariate, USEBOD, to confirm the shape of the relationship (Figure 2). The scatter plot shows a clear linear decreasing pattern of σ_j as USEBOD increases.

At this point, d_1 in equation (3) captures the relationship between the adjusted school mean (β_{0j}) and within-school residual standard deviation (σ). A negative d_1 estimate indicates that schools with high average achievement also have smaller variance, holding constant USEBOD. This could possibly occur due to a ceiling effect or other successful instructional factors. Note that β_{0j} is centered around its grand mean (γ_{00}) so that the intercept (d_0) can represent the average within-school variation of 143 schools. d_2 in equation (3) shows whether teachers' frequent use of the board can reduce σ, after controlling for school mean achievement. USEBOD is also grand mean centered. Results for this variance model (Models 4 and 5) are summarized in Table 5. Since common software packages tailored to multilevel analysis do not provide solutions for latent variable regression for variance, we used a fully Bayesian approach via Markov chain Monte Carlo (MCMC) method (e.g., Gibbs sampler) implemented in WinBUGS (Spiegelhalter, Thomas, Bets, & Lunn, 2003).

Results

The results for Models 1 through 4 are preliminary analyses showing our procedures step-by- step. Therefore, we will discuss only the results for the final model (Model 5). General fixed effects in the mean structure model (achievement model) will be discussed first; then, more importantly, the result for the variance model (dispersion model) will be discussed.

Achievement Model

After adjusting for the effect of parents' education levels, home educational resources, and extra outside school

Table 5
Result Summary for Heterogeneous Variance Modeling

	Model 4			Model 5		
	Mean	95% interval	Prob. >0	Mean	95% interval	Prob. >0
Mean model						
For adjusted grand mean, β_{0j}						
Adjusted grand mean, γ_{00}	591.0	588.4, 593.7	1.000	591.0	588.3, 593.7	1.000
Urban schools, γ_{01}	18.08	7.55, 28.38	1.000	18.26	7.69, 28.65	1.000
Suburban schools, γ_{02}	14.27	3.80, 24.45	0.996	14.55	4.18, 24.78	0.997
School average parent ed. Level, γ_{03}	15.45	5.87, 25.12	0.999	15.28	5.63, 24.96	0.999
Gender contrast, γ_{10}	-2.60	-8.74, 3.55	0.203	-2.44	-8.57, 3.68	0.217
Parent highest ed. Slope, γ_{20}	6.15	3.79, 8.49	1.000	6.16	3.82, 8.51	1.000
Home resource slope, γ_{30}	31.84	26.71, 36.95	1.000	31.89	26.78, 37.00	1.000
For extra outside lesson slope, β_{4j}						
Adjusted average effect, γ_{40}	21.15	16.51, 25.75	1.000	21.01	16.42, 25.60	1.000
Urban schools, γ_{41}	-10.99	-30.02, 7.53	0.121	-10.86	-30.03, 7.99	0.128
Suburban schools, γ_{42}	-4.99	-23.54, 13.46	0.299	-4.81	-23.60, 14.10	0.305
School average parent ed. Level, γ_{43}	-15.33	-31.01, 0.49	0.029	-14.94	-30.69, 0.85	0.031
For 'AO frequency' slope, β_{5j}						
Adjusted average effect, γ_{50}	14.61	12.01, 17.24	1.000	14.63	12.02, 17.25	1.000
Urban schools, γ_{51}	2.32	-7.71, 12.63	0.669	2.73	-7.19, 12.67	0.706
Suburban schools, γ_{52}	2.27	-7.50, 12.38	0.666	2.74	-7.15, 12.61	0.705
School average parent ed. Level, γ_{53}	-10.53	-19.59, -1.49	0.011	-10.68	-19.76, -1.64	0.011
Student use board' slope, γ_{60}	6.63	4.61, 8.67	1.000	6.65	4.61, 8.69	1.000
Positive attitude toward math slope, γ_{70}	27.72	24.52, 30.92	1.000	27.74	24.55, 30.94	1.000
For 'play time' slope, β_{8j}						
Adjusted average effect, γ_{80}	-13.41	-17.65, -9.19	0.000	-13.43	-17.74, -9.14	0.000
Urban schools, γ_{81}	-13.71	-30.11, 3.70	0.056	-13.95	-30.63, 2.35	0.047
Suburban schools, γ_{82}	-13.70	-30.10, 3.30	0.054	-13.76	-30.26, 2.24	0.048
School average parent ed. Level, γ_{83}	3.81	-10.83, 18.58	0.693	4.15	-10.19, 18.90	0.712
Variance model for σ_j						
Average within-school *SD*, d_0	65.62 (0.74)	64.18, 67.10	1.000	65.62 (0.73)	64.21, 67.09	1.000
School mean achievement slope, d_1	-0.14 (0.06)	-0.27, -0.02	0.012	-0.13 (0.06)	-0.25, -0.01	0.019
Teacher ues board slope, d_2				-8.57 (4.31)	-16.92, -0.05	0.024
Random effects variance matrix (**T**)	$\begin{pmatrix} 1403 & -57.73 & -7.20 & -14.73 \\ & 2087 & -12.43 & 35.30 \\ & & 65.32 & -11.06 \\ & & & 1332 \end{pmatrix}$			$\begin{pmatrix} 1398 & -54.38 & -8.68 & -13.23 \\ & 2025 & -11.25 & 35.72 \\ & & 67.69 & -10.82 \\ & & & 1327 \end{pmatrix}$		

math lessons, the grand mean math achievement score is equal to 591 (γ_{00}). The mean for urban schools was 18.26 points above average (γ_{01}). The mean for suburban schools was 14.55 points above average (605.55). The contextual effect of the aggregate parent education level was 15.28 (γ_{03}). Because the standard deviation of MPED is .32 (see Table 2), if we compare two students with the same parental education level in two schools differing by 1 *SD* MPED level, we would expect the student in the school in the 1 *SD* higher MPED level to show 4.89 points (i.e., 15.28×.32≈4.89) higher achievement than the other student in the other school.

We found no gender difference in math achievement (γ_{10}). Parents' education level did make a difference in student math achievement (γ_{20}). Note that the possible difference in math achievement between students in the lowest parents' education level and the highest is 18.48 (i.e., 6.16×3=18.48). However, as mentioned above, depending on the school's average PED level, this gap can get wider or narrower. Home resources had a strong effect on math achievement (the effect estimate is 31.89, γ_{30}). Because this variable is coded 0 to 2, the expected difference between students with low and high home resources is 63.78 (i.e., 31.89×2=63.78). However, note that most of the students (80%, Table 1) had a medium home resources level.

Students who took extra math lessons outside school more than 1 hour per week scored about 21 points higher on average (γ_{40}). However, students in high MPED level schools got less benefit from extra lessons (γ_{43}). Students' frequent exposure to teacher's advance organizer (AO) did increase students' achievement (γ_{50}). In addition, in high MPED level schools, this AO effect was smaller than average (γ_{53}). For example, the average AO effect was 14.63, and the AO effect for schools at 2 *SD* above the average MPED level was 7.79 (14.63 - (2×.32×10.68)≈7.79). The reason for extra lessons and AO being less effective in high SES schools requires further investigation. However, one possible explanation might be that in high SES schools, students could have access to various alternative educational resources and different environments (e.g., peer/family pressure and better classroom instruction) not specified in this study may contribute to student achievement, compared to low SES schools where students have, for example, fewer options to take extra lessons.

Students' more frequent board use was positively associated with math achievement (γ_{60}). Moreover, students reporting a high positive attitude toward math showed a higher degree of math achievement (γ_{70}). These effects did not vary significantly across schools.

γ_{80} indicates that students who spend more than 3 hours doing non-academic activities after school scored 13.43 points less on average. Interestingly, this gap gets wider in both urban (γ_{81}) and suburban schools (γ_{82}). In urban schools, the gap became 27.38 (-13.43 - 13.95 = -27.38), and in suburban schools, the gap is 27.19 (-13.43 - 13.76 = -27.19). In general, the gap between the two activity groups was smaller in rural schools than in non rural schools.

Dispersion Model

Variance model results (d_0 to d_2 in Model 5; see Table 5) show that average σ_j was equal to 65.62 (d_0). School mean achievement was significantly related to smaller σ_j (d_1 = -.13). d_2, the effect of USEBOD, was -8.57 with prob.(d_2 > 0) equal to 0.024. This shows that 97.6% of the posterior distribution of d_2 falls below zero—strong evidence of a negative relationship. Therefore, data supports the idea that even after adjusting for school mean, using the board frequently in classroom instruction seems to reduce the achievement gap within schools. Table 2 shows that teachers already used the board frequently in the classroom (mean=3.05, *SD*=.17). We expect a 1.46 point (8.57×.17≈1.46) decrease in σ_j when USEBOD increases by 1 *SD*. If we compare two schools with a 2 *SD* difference in USEBOD and the same achievement level, the school with higher USEBOD will have about a 11.4 points smaller 95% interval.[3] This interval can alternatively be interpreted as the gap between the upper and lower 2.5% achievement level in a school. Therefore, the gap between the upper and lower 2.5% students will also be smaller by 11.4 points in schools with 2 *SD* above the USEBOD level. This variance model result is summarized in Figure 3. Each bar in Figure 3 represents the predicted 95% achievement range in a school.

As noted before, the 95% achievement range can be interpreted as the achievement gap between the highest and lowest 2.5% of students in a school. Figure 3 presents a graphical representation of the results from the dispersion modeling in Model 5. We present three schools with different achievement levels (2 *SD* below average, average, and 2 *SD* above average), and within each achievement

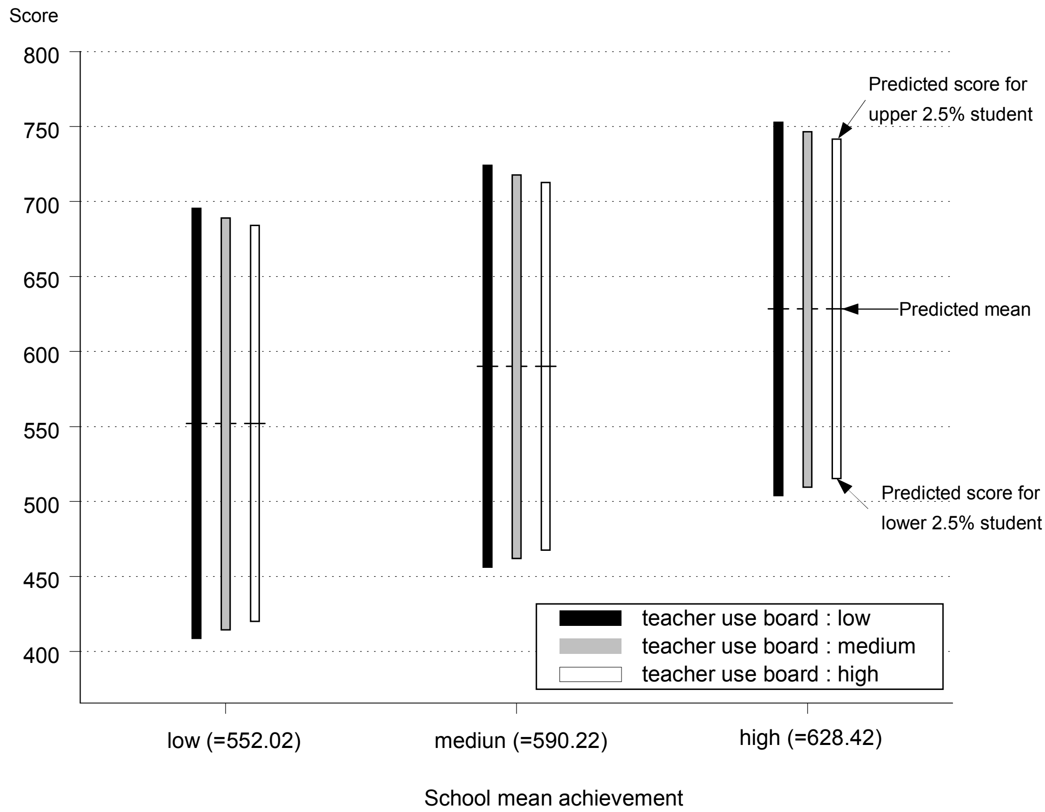

Figure 3. Comparing achievement gap between high and low achieving students in schools with different achievement level and USEBOD level.

level, we selected three USEBOD levels (2 *SD* below average, average, and 2 *SD* above average). This figure shows that high-achievement schools have a smaller gap, and among schools with the same achievement level, schools with high USEBOD schools tend to have a smaller gap. However, we recommend that readers interpret the result with caution because USEBOD is chosen solely for illustrative purposes and is not based on a systematic theory. For example, it is likely that USEBOD works as a proxy of the underlying construct or school practice. Clearly, further research is needed to identify factors that are associated with homogenizing the competency levels of students in a school.

Summary and Discussion

In this study, we tried to answer some important questions in school effect studies, such as: 'what elements make a good school?' and 'what kind of school is effective in closing the gap between high- and low-achieving students?' In this regard, we argue that effective schools not only increase student achievement levels on average, but also reduce the gap between student achievement levels. Through the illustration with data from TIMSS-R, we showed how simultaneous modeling of within-school variation and achievement levels under a multilevel framework can be utilized to answer these questions – that is, to identify effective schools and the correlates of a smaller achievement gap. Our analysis showed that schools did vary, not only in their achievement levels but also in the achievement gap between high- and low-achieving students.

To illustrate examples of more and less effective schools, we chose three schools in Figure 4, based on our results. Note that solid reference lines represent the estimated upper 2.5% achievement level, the grand mean, and the lower 2.5% achievement level in the population,

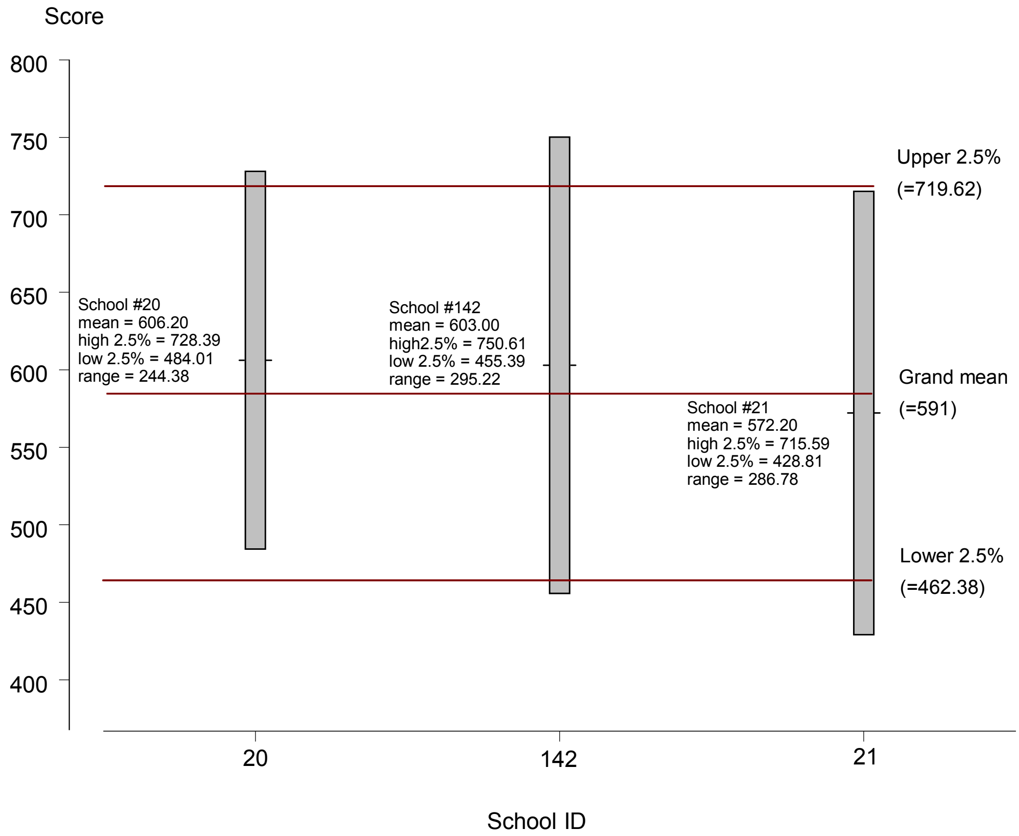

Figure 4. Contrasting three type of schools: High achievement and small gap (#20), High achievement and large gap (#142) and Low achievement and large gap (#21).

respectively.

First, schools #20 and #142 have similar mean achievement levels (606 and 603). However, if we compare the predicted gap between the upper and lower 2.5% of students in the two schools, we see that the gap is about 50 points smaller in school #20. Therefore, in terms of closing the gap, school #20 is more effective than school #142. School #21 is an example of a less effective school in terms of dispersion as well as achievement, that is, low achievement and a larger gap. As shown, one advantage of proposed variance modeling is that we can actually calculate the gap between any two achievement percentile scores within a school (for example, 25% and 75%), and this can be used as a school indicator along with school performance level.

Next, our proposed analytic method enables us to study the school characteristics or practices that reduce or magnify the gap, which can provide information for school reform to direct as many students as possible towards the achievement goal. Using the latent variable regression technique, we modeled the latent school-specific variance heterogeneity as a function of school-level observed (USEBOD) and latent (average achievement) variables. Incorporating latent average achievement levels in the dispersion model is especially useful because, by blocking by achievement level, one can effectively control for the ceiling effect on the inference regarding the effect of school practice on the achievement gap.

In simultaneous modeling of dispersion and mean structure, the specification of mean structure is especially important because the detected difference in dispersion among schools with different levels of certain school characteristic may signal unnoticed interaction between the school characteristic and certain student characteristics, whether observed or not (Kim & Seltzer, 2006). Checking whether the specified mean structure adequately represents the data or not is especially challenging in complex models, and has not been studied sufficiently in educational research. It is also possible that unnoticed student variables may affect the inference in terms of the dispersion and subsequent studies are needed to provide the basis for rigorous discussion as to why some schools are more homogeneous in achievement than others. One way which shows promise for assessing how well the model represents the observed data is to simulate data from the fitted model and compare it with the observed data, a technique called Bayesian posterior predictive model checking. Readers are referred to Gelman et al. (1995) and Kim and Seltzer (2008) for detailed discussion on this topic.

Notes

[1] To examine the distribution of σ^2_j, we first fit the achievement model specified in equations (1) and (2) and obtained the estimate of σ^2_j, assuming that each school has its own level-1 variance.

[2] The result for Model 5 without school #142 is as follows: d_0 = 65.37 (*prob*.(d_0> 0) = 1.00), d_1 = -1.83 (*prob*(d_1> 0) = 0.26) and d_2 = -8.15 (prob.(d_2> 0) = 0.01).

[3] The 95% interval, which captures the middle 95% of the predicted achievement distribution in a school, can be calculated as $\beta_{0j} \pm 1.96\sigma_j$. This interval becomes smaller in schools with high achievement or with higher USEBOD level, because σ_j becomes smaller in these schools.

References

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, *104*, 396-404.

Choi, K., & Seltzer, M. (in press). Modeling heterogeneity in relationships between initial status and rates of change: treating latent variable regression coefficients as random coefficients in a three-level hierarchical model. *Journal of Educational and Behavioral Statistics.*

Eugenio, J. G., & Julie, A. M. (2001). *TIMSS 1999 user guide for the international database*. Boston: Boston College, Lynch School of Education, International Study Center.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (1995). *Bayesian data analysis*. New York, NY: Chapman & Hall.

Kasim, R., & Raudenbush, S. W. (1998). Application of Gibbs sampling to nested variance components models with heterogeneous within-group variance. *Journal of Educational and Behavioral Statistics*, *23*, 93-116.

Kim, J., & Seltzer, M. (2008). *Bayesian model checks for complex hierarchical models in quasi-experimental settings*. Paper presented at the 2008 annual meeting of American Educational Research Association.

Kim, J., & Seltzer, M. (2006). *Examining heterogeneity in residual variance in experimental and quasi-experimental settings*. Paper presented at the 2006 annual meeting of American Educational Research Association.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62,* 172-192.

Park, D., Park, J., & Kim, S. (2001). The effects of school and student background variables on math and science achievements in middle schools. *Journal of Educational Evaluation*, *14*, 127-149.

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

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *HLM6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Rumberger, R. W., & Palardy, G. J. (2003). Multilevel models for school effectiveness research. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 235-258). Thousand Oaks, CA: Sage.

Seltzer, M., Choi, K., & Thum, Y. (2003). Examining relationship between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Education Evaluation and Policy Analysis*, *25*, 263-286.

Spiegelhalter, D., Thomas, A., Bets, N., & Lunn, D. (2003). *WinBUGS: windows version of Bayesian inference using Gibbs sampling, version 1.4, User manual*.

Cambridge, UK: University of Cambridge, MRC Biostatistics Unit.
Yang, J., & Kim, K. (2003). Effects of middle school organization on academic achievement in Korea: An HLM analysis of TIMSS-R. *Korean Journal of Sociology of Education*, *13,* 165-184.

Received December 10, 2007
Revision received February 23, 2008
Accepted March 14, 2008