Classroom and Teacher Contextual Effects on Students’ Science Concept Learning: A Multilevel Analysis

Vasuki Rethinam
Curtis Pyke
&
Sharon Lynch

Abstract

This paper builds on our prior work in applying HLM analysis techniques to multilevel curriculum evaluation data by modeling additional classroom and teacher variables as predictors of students’ learning in a middle school science curriculum experiment. The sample consisted of grade 6 students (n = 4577, classroom n = 183, teacher n = 45) from a large metropolitan school district. Null model showed that 17% of variance exists between teachers indicating teachers matter in a curriculum evaluation study. The full model explained 24% of the variance in students’ outcome. Experimental condition was significantly related to outcome after controlling for teacher effects. Student level variables still matter even after controlling for class- and teacher-level variables in a 3-level HLM model.

Objective/ Purpose

The purpose of this paper is to extend our prior work in applying HLM to quasi-experimental curriculum evaluation data that is nested. We use data from schools implementing the science curriculum unit at scale. We examine the effectiveness of curriculum unit after controlling for individual, classroom and teacher effects.

Background

This work is part of a NSF funded project evaluating curriculum to better understand the effectiveness of reform-based curriculum materials as a precursor to their scaling-up in a large and diverse school district. The evaluation research questions studied by the project are: “does the new curriculum work?” and “does the new curriculum close gaps?” In the project’s studies students were not randomly assigned to a curriculum condition and so students were “naturally” nested within classrooms and schools. The project has employed ANCOVA techniques with posttest scores as the outcome and pretest scores as covariate to test hypotheses that differences in outcomes result from using the new curriculum (Lynch, Kuipers, Pyke, & Szesze, 2005; Lynch, Taymans, Watson, Ochsendorf, Pyke, & Szesze, 2007). In our prior work we applied HLM analyses to the projects data because of its potential to address concerns about liberal significance tests derived from ANCOVA analyses with nested data.

The nesting of the project’s data is a statistical concern because context can affect all students grouped in a classroom and confound observation of a curriculum’s effect. When we reanalyzed data from the project’s Exploring Motion and Forces (M&F)
HLM analysis was also conducted on another set of third year implementation of M&F in a set of 5 matched pairs of new schools. This analysis used an improved model of classroom context such as low- and high-achievement history classrooms. The results of this analysis also indicated that the classroom context was important. We found a cross level interaction between ESOL students and low achievement history and Hispanic students and high achievement history classrooms. In other words, ESOL students in low achievement history classrooms were at a disadvantage compared to ESOL students in the medium achievement history classrooms. Hispanic students in high achievement history classrooms were at an advantage compared to Hispanic students in medium achievement history classrooms. This analysis further confirmed the importance of context effects in curriculum evaluation research (Authors, 2007).

The current paper uses a new set of data to examine the effectiveness of curriculum unit at scale. Here we employ an improved model of context by extending the model to a 3 level by including teacher variables at the highest level. Our goal was to better account for the curriculum effects by controlling for more known variability and accounting for more variation overall.

**Perspective/Theoretical Framework**

In this paper, we continue to develop a theoretical and practical rationale for multilevel modeling of classroom and teacher effects. We argue that it is not enough to look only at individual student characteristics to explain the success or failure of an intervention. It is also important to understand the contexts in which interventions are implemented and learning occurs (McDonald, Keesler, Kauffman, & Schneider, 2006). Classroom and teacher contextual effects can influence learning because each student is affected by multiple factors, widely understood to account for variation in learning. There is empirical evidence from sociological and psychological studies of school- and classroom-level influences that classroom conditions do impede or promote learning (see Ames & Ames, 1984 and McDonald et al., 2006).

Assignment of students to classrooms within schools and between schools is an important contextual factor that has not always been examined in curriculum effectiveness studies, although prior studies indicated that instruction in high- and low-track classrooms produces different student outcomes (Callahan, 2005; Van Houtte, 2004). However, in most prior studies, tracking/ability grouping variables were used as school level variables and data were analyzed using single level regression analyses (Carbonaro, 2005; Marks 2006). Other studies that used multilevel analysis included the tracking/ability grouping variable at the school level (McCoach, O’Connell, & Levitt,
2006; Van Houtte, 2004) but have ignored the classroom level analysis of these variables. Our prior HLM analysis included the achievement history variable at the classroom and level in a two-level analysis. Since the results showed significant relationship between achievement history and science learning, we include this variable at the classroom level in this paper.

We add to the growing research on classroom context effects here by adding to our model another level – the teacher effects on students’ science learning. Studies have indicated that teachers have a significant influence on student learning (Darling-Hammond, & Young, 2002; Odden, Borman, & Fermanich, 2004). Based on a literature review, Odden et al. (2004) identified some teacher factors that were found to be associated with student achievement. Some of these factors include: Years of teaching (Goldhaber & Brewer, 1997), major of undergraduate study, particularly for mathematics and science teachers (Monk, 1994), course work or degree obtained (Rowan, Chiang, & Miller, 1996), and earning of license (Darling-Hammond & Young, 2002). In this paper, we use the available teacher variables in our dataset: Years of teaching science and science credit hours taken. We also included “years of implementing M&F.” Our expectation is that these above teacher variables will influence outcomes. The intent of introducing these teacher variables is to better control for teacher effects to provide a more precise estimate of curriculum effects and to continue to explore context effects by including a third level (teacher variables) in a multilevel analysis.

Research Questions

1. Do teachers matter?
2. Does curriculum condition still matter after controlling for class level (achievement history), teacher level factors (Science credits taken, years taught science), and individual level factors?
3. Does the curriculum condition interact with individual level demographic variables?
4. Do student level variables still matter after controlling for classroom and teacher effects?

Method

Sample

The sample of the study included grade 6 (n = 4577, classroom n = 183, teacher n = 45) students from 20 middle schools each selected from within a large northeastern metropolitan school district. Of the 20 schools, 5 schools implementing M&F for the 1st time and 5 comparison schools not implementing M&F was selected. The ten schools were matched on demographic variables resulting in five matched-pairs of schools. One school from each pair was randomly selected to implement the highly rated M&F curriculum unit. The comparison group experienced the regular curriculum offered by the school district. The remaining 10 schools from 2 matched sets of schools, each set of schools implementing M&F at a different stage of scaling, were included.
Variables

Student level. FARMS, SPED, ESOL (dichotomous variables), dummy coded ethnicity (Asian American, African American, and Hispanic - white was used as a reference group), and MFA pretest score.

Classroom level. Experimental condition (dichotomous) and classroom achievement history were used. The following procedure was used to create the classroom achievement history variable. Classrooms within schools were rank-ordered based on the means of three student-level criteria, (a) prior science GPA in the same academic year, (b) standardized math test scores, and (c) standardized reading test scores obtained from 5th grade tests. The ranked scores for the three criteria were summed to provide two highest achievement classrooms and two lowest achievement classrooms for each school. Classrooms with scores in the middle were coded as a third category (reference category, or medium achievement history). This procedure was conducted for each school. Thus, the class mean achievement history is a naturally occurring classroom composition variable within schools. This variable, mean achievement history has three levels (high, medium, and low) within each school.

Teacher level. A factor analysis was conducted using the teacher variables available in the database. The analysis produced a two factor structure: science credits taken by the teacher and number of years teaching science. We also included “years of implementing M&F.”

Outcome variable. Student posttest scores. MFA consists of 10 items that require understanding of forces and motion. Student scores on MFA were transformed into weighted scale scores ranging from 0 to 100.

Data Analyses

Using the HLM program (Raudenbusch, Bryk, Cheong, & Congdon, 2004), hierarchical linear modeling technique was employed for the data analyses. The continuous variables modeled in this analyses were converted to a z-score (M = 0, SD = 1). The categorical variables were dummy coded as 0 and 1.

A three level HLM analysis were conducted. HLM analysis consists of three stages for this study. The first is a fully unconditional model. The second stage of HLM employed is referred to here as the “intercept as outcomes model,” as indicated by Raudenbush and Bryk (2002). The third and final stage of the HLM employed a cross-level interaction, that is, an interaction between classroom level curriculum condition and individual level student demographic variables.

Results

Unconditional Model
The unconditional model estimations indicate that the level 1 model, which explains the proportion of variance in the outcome that exists between students accounted for 78% of the variance; level-2 model, which explains the proportion of variance in outcome that exists between classrooms, accounted for only 5% of the variance and level-3 model, which explains the proportion of variance in outcome that exists between teachers, accounted for 17% of the variance.

Final Model

We present effect sizes of the significant variables. At the between teacher (level 3), none of the teacher variables were found to be significant even though the null model indicated that 17% of variance could be attributed to the outcome. At the classroom level (level 2), only experimental condition was found to be significant. Comparison classrooms scored .66 SD points more than treatment classrooms. Surprisingly, none of the classroom achievement history variables were significant. At the student level: Students’ SPED status had a significant influence on their posttest scores (ES = -.92 SD, p<.0001); FARMS (ES = -.39 SD, p<.0001) and ESOL (ES = -.39, p<.0001) status were significantly inversely related to student posttest scores. African American (ES = -.49 SD, p < .001) and Hispanic (ES = -.27 SD, p < .01) students variables were significantly related to students’ posttest scores. Students with greater scores on pretest scored .76 SD points more on posttest than students with lower pretest scores. In addition, two significant cross-level interactions were found between FARMS and experimental condition; and SPED and experimental condition.

Interpretations/Conclusion

The full model explained 24% of variance in posttest scores. The null indicated that the proportion of variance that existed between teachers were 17% thereby suggesting that teachers do matter. Of interest is that taking into account all other variables, there was a significant difference found between treatment and comparison condition on posttest scores. Comparison classrooms scored higher than the treatment classrooms. There were significant cross-level interaction between curriculum condition and individual level demographic factors. It was interesting to note that the student level variables remained significant after controlling for classroom and teacher level variables. Students eligible for SPED & FARMS services, students eligible for ESOL, African American and Hispanic students were scoring lower than their peers on posttest scores. Individual level variables still matter even after controlling for class- and teacher-level variables.

Educational Importance

The purpose of this paper is to extend our prior work in applying HLM to quasi-experimental curriculum evaluation data that is nested. We use data from schools implementing the science curriculum unit at scale. The current paper improves upon the prior HLM model to account for more explained variance in the outcome by including a third level (Teacher variation). School personnel should be concerned about students eligible for SPED, ESOL and FARMS services, and Hispanic and African American
students. Teacher variation seems to matter, even though none of the teacher variables were significant. Other teacher variables should be explored. However, it is not appropriate to ignore the context (classroom and teacher) in studies that include nested data.

References


Author3, -----, Author2, -----. (2005).

Author3, -----, ----, -----, Author2, -----. (2007).


Authors. (in press).

Authors. (2007).
