

Research Article

Are School Factors Important for Measuring Teacher Effectiveness? A Multilevel Technique to Predict Student Gains through a Value-Added Approach

Bidya Raj Subedi,¹ Bonnie Swan,² and Michael C. Hynes³

¹ Department of Research and Evaluation, the School District of Palm Beach County, West Palm Beach, FL 33406, USA

² Program Evaluation and Educational Research Group (PEER), College of Education, University of Central Florida, Orlando, FL 32816, USA

³ Department of Mathematics Education, College of Education, University of Central Florida, Orlando, FL 32816, USA

Correspondence should be addressed to Bidya Raj Subedi, bidya.subedi@palmbeachschools.org

Received 18 November 2010; Revised 17 February 2011; Accepted 23 March 2011

Academic Editor: Stephen P. Heyneman

Copyright © 2011 Bidya Raj Subedi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper investigated the effect of teacher quality, represented by teacher level characteristics, on mathematics gain scores employing a three-level hierarchical linear model (HLM) through value-added model (VAM) approach. The analysis investigated significant predictors at student, teacher, and school levels for predicting students' gain scores and also estimated d-type effect sizes at teacher and school levels. We found the significant effects of teacher's mathematics content certification, teacher experience, and the interaction effects of mathematics content certification with student level predictors. Although school poverty significantly predicted students' gain scores, the school level effect was relatively small.

1. Introduction

Student achievement gain can be predicted due to individual predictors at student, teacher, and school levels. This predictive model generates a multilevel research design, known as hierarchical linear model (HLM), which also allows the cross-level interactions among all possible predictors at student, teacher, and school levels. Employing three-level HLM through value-added approach, this paper aims to measure the teacher quality, represented by teacher effectiveness, based on three teacher level factors: teacher content-area certification, teacher experience and teacher's advanced mathematics or mathematics education degree. Individual teachers in Florida schools must be formally certified in the content area that they teach. To earn this certification, they must hold at least a bachelor's degree plus demonstrate mastery of subject area knowledge or meet subject specialization. Although having a degree in mathematics or mathematics education makes it easier

to demonstrate subject area knowledge or meet subject specialization, most mathematics teachers come through a different route.

This study used middle school mathematics data from the Orange County Public Schools (OCPS), which is the fourth largest school district in Florida and tenth largest in the USA based on student enrollment. According to the Florida Department of Education (FLDOE) 2010 report, this school district has an enrollment of 175,986 from prekindergarten to grade 12 with 194 schools in the district. Out of total enrollment, this district has 32% White, 27% Black, 33% Hispanic students compared to 43% White, 23% Black, and 28% Hispanic students in Florida. The poverty data published from FLDOE show that this district has 57% students participating in free and reduced lunch (FRL) program compared to 56% students in the state participating in FRL program. According to FLDOE report, 59% of the teachers in OCPS have bachelor's degree and 41% have master's level and above degrees; such result is similar to

the percents of teachers in the state of Florida holding bachelor's and advanced degrees. The average years of teacher experience in this district and Florida is 12 years. The average yearly salary of OCPS teachers is \$44,884 that is little lower than the average yearly salary of teachers in the state of Florida, which is \$46,696.

In this paper, we examine the effects of student, teacher, and school level factors, including cross-level interaction effects, on students' mathematics gain scores. These gain scores, which are different from teachers' value-added scores, are computed as the difference between students' current and prior years' scores obtained from the state mandated standardized test. We also report the variances explained and effect sizes at school and teacher levels in order to measure teacher effects by predicting students' mathematics gain scores.

2. Review of Studies in Pertinent Area

2.1. Value-Added Models and Decomposition of Variances for Teacher Effectiveness. Value-added models (VAMs) measure teacher's contribution to improving students' gain scores for the same cohort of students in a given school year. The VAMs seek to isolate the contribution that each teacher makes in a given school year, which can be compared to the performance measures of other teachers. According to Rowan, Correnti, and Miller [1], the purpose of value-added models is to estimate the proportions of variance in changes in student achievement among classrooms, after controlling for the effects of other confounding variables. Tobe [2] mentioned that the differences between teachers can be quantified as "teacher effects" using value-added models. Because VAM studies can show very large differences in effectiveness among teachers, the potential gains in academic achievement can be measured if these differences are substantiated and causally linked to specific teacher characteristics. Other powerful value-added models (e.g., [3, 4]), that track students' gains over more than one year, have brought about a rethinking among researchers regarding the relative importance of the role of teachers. Sanders and Rivers' ground-breaking Tennessee value-added study showed that fifth grade mathematics students matched in performance assigned to ineffective teachers for three years performed dramatically worse (separated by 50 percentile points on comparable assessments) than children assigned to more effective teachers. Similarly, Jordan, Mendro, and Weerasinghe [3], who isolated the effects of Texas teachers on student achievement, found differences of 34 percentile points in reading and 49 percentile points in mathematics achievement, when comparing students assigned to ineffective teachers for three consecutive years to students assigned for three years to effective teachers (defined by how much their students improved). Since there are no study in the past employing three-level HLM through VAM approach with Orange County middle school mathematics data, this study fulfils the research gap in terms of determining the significant factors at student, teacher, and school levels for measuring teacher effectiveness using this new approach.

Recent studies have addressed the relationship between student achievement as well as gains and the factors at student, teacher, and school levels by using variance decomposition approach in HLM. For example, in a review of multilevel studies relating to teacher quality and student achievement, Scheerens and Bosker [5] found that the differences in student achievement are associated with school (20%) and classroom/teacher level factors (20%), with the remaining difference (60%) at the student level factors (such as socioeconomic status and prior achievement). Rowan et al. [1] employed three-level HLM incorporating predictors at student, classroom, and school levels in order to predict mathematics and reading achievement and annual gains. They allowed variance decomposition among students, classrooms, and schools in order to measure teacher effectiveness. However, little or no studies have been conducted so far in the pertinent area using variance decomposition method employing three-level HLM with Orange County data. Therefore, this study uses such method to predict the middle school mathematics gain scores of Orange County providing explained variances separately at teacher and school levels in order to determine teacher effectiveness.

2.2. Determining the Factors of Teacher Quality and Reporting d-Type Effect Size. In the context of mandates and philosophies of the No Child Left Behind (NCLB) Act in United States of America (USA), much of what is driving educational reform centers on the premise that a teacher's background matters. For example, by the end of the school year 2005-2006, for the first time, states were required to have data collection and reporting mechanisms in place to publish reports disclosing whether they met the goal of ensuring all teachers are "highly qualified." Meeting these standards basically means that teachers must (a) hold an acceptable bachelor's or higher degree, (b) have state licensure or certification, and (c) demonstrate subject competency of the subject(s) at the grade level(s) taught.

Past research has evidenced that the need for qualified teachers is particularly great in lower-performing schools with higher numbers of low-income and minority students (see [4, 6-10]), and the problem is even more pronounced in middle schools (see [11]).

Evidence is mounting that better teachers can and do make a difference in student achievement [3, 4, 12]. Still, substantial disagreement exists among researchers as to what teacher qualifications make a difference [13] and little has been explored on this topic specific to the middle school classrooms. Further, Rice [14] found a serious gap in the knowledge base that still needs to be explored regarding middle schools (and elementary schools) teachers' effectiveness that is used to guide important teacher policy decisions. Her award-winning review examines the impact of teacher characteristics on teacher effectiveness. In a study related to eighth grade students' mathematics achievement using 1996 National Assessment of Educational Progress (NAEP) data, Wenglinsky [15] found that the effects of classroom practices, when added to those of other teacher characteristics, are comparable in size to those of student backgrounds, suggesting that teachers can contribute as much to

student learning in mathematics as the students themselves. Through a research on teacher qualification, Croninger, Rice, Rathbun, and Nishio [16] found potential contextual effects of teachers' qualifications on student achievement, with first graders demonstrating higher levels of reading and mathematics achievement in schools where teachers report higher levels of coursework emphasis in these subject areas.

Darling-Hammond, Holtzman, Gatlin, and Heilig [17] found that certified teachers consistently produce stronger student achievement gains than do uncertified teachers and controlling for teacher experience, degrees, and student characteristics, uncertified teachers are less effective than certified teachers. Darling-Hammond [18] found that measures of teacher preparation and certification are by far the strongest correlates of student achievement in reading and mathematics, both before and after controlling for student poverty and language status. Decker, Mayer, and Glazerman [19] found that teachers recruited through Teach for America (TFA) are significantly more effective than both uncertified and certified teachers at mathematics instruction and statistically indistinguishable in reading instruction. However, Kane, Rockoff, and Staiger [20] found no difference between teaching fellows and certified teachers or between uncertified and certified teachers in their impact on mathematics achievement. Subject content-area certification has a major role in significantly impacting student achievement. For example, Goldhaber and Brewer [21] found that mathematics teachers who have a standard certification have a statistically significant positive impact on student test scores relative to teachers who either hold private school certification or are not certified in their subject area.

Relevant studies address the effect of teacher degree and experience on student mathematics achievement (see [14, 18, 22–27]). In past, Swan and Subedi [28] and Swan [27] previously found far fewer teachers with advanced degrees teaching in poor schools as opposed to the number of similar teachers instructing in wealthier schools. Since the past studies are insufficient in determining the factors of teacher quality in the context of population under study, this paper explores whether factors representing teacher quality are significant in predicting students' mathematics gain scores in middle schools.

Researchers, in the past, have used student level predictors in multilevel model by incorporating students' prior achievement and socioeconomic background in the model to predict mathematics and reading achievement (see [1, 5]) including cross-level interactions with predictors at higher level (see [29]). Through value-added model to measure teacher effects, Rowan et al. [1] used students' prior achievement, socioeconomic status, and school poverty to predict students' gain scores employing three-level HLM. Further, Banks [30] investigated the effect of school poverty concentration on student achievement.

The common problems in educational research are to explore the effect sizes at teacher and school levels by identifying significant teacher and school level predictors. The educational research and evaluation area demands

these kinds of studies particularly in large urban school districts, such as Orange County, so that implications and generalizations can be made based on valid results. With the aim of practical significance in addressing contemporary issues, this research predicts students' gain scores due to potential factors at student, teacher, and school level models in order to measure teacher effectiveness, given by d-type effect size, through VAM employing three-level HLM.

3. Methods

3.1. Design of the Study. This study explores the individual effects of teacher level predictors: mathematics content-area certification, advanced mathematics or mathematics education degree, and experience. Subedi and Swan [31] conducted similar study using two-level HLM analysis. This study extends their work using three-level HLM through value-added model in order to measure teacher effectiveness. Since the students were not assigned randomly within teachers' classrooms and the predictors incorporated separately in student, teacher, and school level models provide better estimates of variance and predictors' effects, the most appropriate statistical design to measure the teacher effects involves a multilevel or an HLM technique (see [32–34]).

3.2. Data and Variables. This study used 6,184 students and 253 mathematics teachers from all middle schools in the Orange County Public Schools (OCPS), which is the tenth largest school district out of 14,000 in the USA.

3.3. Outcome variable. We used grades 6–8 mathematics gain scores as an outcome variable. The gain scores are calculated as the difference in scores of 2005 and 2004 NRT-NCE (Norm Referenced Test-Normal Curve Equivalent) portion of the FCAT (Florida Comprehensive Assessment Test). (The FCAT is a state mandated standardized test which measures student achievement of the benchmarks in reading, mathematics, science, and social studies in Florida schools. The FCAT also provides feedback and accountability indicators to Florida educators, students, parents, and policy makers.) The NRT-NCE scores for this study ranged from 1 to 99 and the gain scores ranged from –31.4 to 45. At the time of the study, the test measured the same mathematics learning benchmarks across grades 6–8 in middle schools throughout Florida.

3.4. Student Level Predictors. The student level predictors used in this study were pretest scores (i.e., NRT-NCE scores for 2004) and student socioeconomic status (SES). Student socioeconomic status (SES) was coded 1 for participation and 0 for nonparticipation in the free and reduced lunch program. Since the data related to student's family income and parental education levels as well as occupations were not available, the indicator of SES is limited to the student's participation status in free and reduced lunch program.

3.5. Teacher Level Predictors. Teacher's content-area certification, a dichotomous predictor, is coded as 1 for in-content certified (indicating holding a mathematics content-area teaching certification for grades 5–9 or 6–12) and 0 for not in-content certified (indicating certification did not qualify them as in-content including elementary education, other education, or other certification). The advanced mathematics degree, another dichotomous predictor, is also coded as 1 for relevant degree (indicating holding an advanced degree in mathematics or mathematics education by a teacher) and 0 for not relevant degree (indicating not holding such a degree). Note that advanced degree is defined as master or higher level degree. Teacher experience, a continuous predictor based on seniority of a teacher, was measured in number of years the teacher taught. This variable ranged from 0 to 37 for this study.

3.6. School Level Predictors. School poverty is defined as the percent of free and reduced lunch students in each school, and teachers' school mean experience is defined as the average number of years taught by middle school teachers in a given school.

3.7. Research Questions. The following research questions are explored through this study:

- (1) what are the significant predictors at student, teacher, and school levels, including cross-level interaction terms, for predicting students' gain scores using conditional VAM?
- (2) what are the proportions of variance explained and effect sizes at teacher level (for measuring teacher effectiveness) and at school level for unconditional model and conditional VAM?

3.8. Why to Use HLM? Hierarchical data structures are present in educational settings where students are nested within a teacher and teachers are nested within a school. The nesting form of data structure generates a hierarchical linear model (HLM). In other words, models at different levels can be built based on specific number of lower level units nested within upper level, eventually forming a HLM design. In the situations where such nesting occurs, the relationship between outcome and predictors can be extended to more than one level, or single level model, such as multiple regression, will not be an appropriate model to use. Thus, in such situations, students' gain scores can be predicted due to the predictors not only at student level but at teacher and school levels.

Since the students are not placed within teachers' classrooms randomly and predictors at student, teacher, and school level models separately provide better estimates of variance and predictors' effects, our best choice of statistical design to estimate the variance and predictors' effects involves selecting the HLM technique. According to many researchers, HLM can be used as an appropriate data analysis method in such situation [32–36].

3.9. Model Development. This study employed a three-level HLM where student, teacher, and school data are incorporated in level-1, level-2, and level-3 models, respectively, to predict students' gain scores. Pretest scores and SES are used as student level predictors at level-1 model. Content-area certification in mathematics, experience, and advanced degree in mathematics or mathematics education are included as teacher level predictors at level-2 model. School poverty and teachers' school mean experience are used as school level predictors at level-3 model. The continuous predictors at level-1 and level-2 models were centered to their grand mean and group mean, respectively. The dichotomous predictors were not centered.

First, level-1, level-2, and level-3 unconditional models, which did not include any predictors at any level, were developed. The proportion of variance explained and effect sizes were calculated at teacher and school level models in order to answer research question 2.

In order to predict mathematics gain scores, the unconditional models at level-1, level-2, and level-3 can be developed as follows:

$$\text{Level-1: } (\text{MATHGAIN})_{ijk} = \pi_{0jk} + e_{ijk},$$

$$\text{Level-2: } \pi_{0jk} = \beta_{00k} + r_{0jk}, \quad (1)$$

$$\text{Level-3: } \beta_{00k} = \gamma_{000} + u_{00k}.$$

In the above equations, π_{0jk} , β_{00k} , and γ_{000} are the intercepts and e_{ijk} , r_{0jk} and u_{00k} are the error terms at student, teacher, and school level models, respectively. We want to estimate r_{0jk} , and u_{00k} to find the proportion of explained variances and effect sizes based on these variances at teacher and school levels, respectively, in order to answer the research question 2.

Since the purpose of value-added model (VAM) is to estimate the proportions of variance in student gain scores lying among teachers after including important predictors in level-1, level-2, and level-3 conditional models, we successively developed such models. The level-2 and level-3 variance terms were deleted from these model if either they were not significant or did not explain more variance in student gain scores after including the error terms in the model. Further, Subedi [34] suggested the formulation of level-2 and level-3 conditional models only after the evidence of significant variance components at level-2 and level-3. In order to predict the mathematics gain scores for student i , taught by teacher j in school k , the level-1 conditional model can be expressed as follows:

$$\begin{aligned} (\text{MATHGAIN})_{ijk} = & \pi_{0jk} + \pi_{1jk} (\text{PRESCORES}) \\ & + \pi_{2jk} (\text{SES}) + e_{ijk}, \end{aligned} \quad (2)$$

where π_{0jk} is the mean gain scores for teacher j in school k , π_{1jk} and π_{2jk} are effects of pretest scores and SES, respectively, at student level, and the term e_{ijk} is the random effect for student i nested within teacher j and school k that is distributed normally with mean 0 and variance σ^2 . Level-2 conditional model for teachers within school can be expressed as

$$\begin{aligned} \pi_{0jk} &= \beta_{00k} + \beta_{01k} \text{ (CERTICON)} \\ &\quad + \beta_{02k} \text{ (TCHREXP)} + r_{0jk}, \\ \pi_{1jk} &= \beta_{10k} + \beta_{11k} \text{ (CERTICON)}, \\ \pi_{2jk} &= \beta_{20k} + \beta_{21k} \text{ (CERTICON)}, \end{aligned} \tag{3}$$

where β_{00k} , β_{10k} , and β_{20k} are the intercepts associated with level-2 model. Further, β_{01k} , β_{02k} , β_{11k} , and β_{21k} are the slopes associated with level-2 model and the term r_{0jk} is the random effect for teacher j nested in school k .

Level-3 model for schools can be given by (4) as follows:

$$\begin{aligned} \beta_{00k} &= \gamma_{000} + \gamma_{001} \text{ (SCHLPOVERTY)} + u_{00k}, \\ \beta_{01k} &= \gamma_{010}, \\ \beta_{02k} &= \gamma_{020}, \\ \beta_{10k} &= \gamma_{100} + \gamma_{101} \text{ (PCTADVMTHDEG)}, \\ \beta_{11k} &= \gamma_{110}, \\ \beta_{20k} &= \gamma_{200} + \gamma_{201} \text{ (MEANEXP)}, \\ \beta_{21k} &= \gamma_{210}. \end{aligned} \tag{4}$$

After substituting (4) in (3) and (3) in (2), the single equation can be expressed as follows:

$$\begin{aligned} &\text{(MATHGAIN)}_{ijk} \\ &= \gamma_{000} + \gamma_{001} \text{ (SCHLPOVERTY)} \\ &\quad + \gamma_{010} \text{ (CERTICON)} \\ &\quad + \gamma_{020} \text{ (TCHREXP)} + \gamma_{100} \text{ (PRESCORES)} \\ &\quad + \gamma_{110} \text{ (PRESCORES * CERTICON)} \\ &\quad + \gamma_{200} \text{ (SES)} \\ &\quad + \gamma_{101} \text{ (PRESCORES * PCTADVMTHDEG)} \\ &\quad + \gamma_{201} \text{ (SES * MEANEXP)} \\ &\quad + \gamma_{210} \text{ (SES * CERTICON)} \\ &\quad + e_{ijk} + r_{0jk} + u_{00k}. \end{aligned} \tag{5}$$

Equation (5) consists of fixed portions (containing γ terms) and random portions (containing e , r , and u terms) of effects. The term γ_{000} represents the grand mean or mean gain scores for all schools and γ_{100} , γ_{200} , γ_{010} , γ_{020} , γ_{001} are the effects of pretest scores, SES, content-area certification, teacher experience, and school poverty, respectively. The factor γ_{110} is the interaction effect between teacher's mathematics content-area certification and students' pretest scores, γ_{101} is the interaction effect between pretest scores and advanced mathematics or mathematics education degree, γ_{201} is the interaction effect between student SES and school teacher mean experience, and γ_{210} is the interaction effect between teacher's mathematics content-area certification and SES. Further, e_{ijk} , r_{0jk} , and u_{00k} are random error terms at student, teacher, and school levels, respectively.

3.10. Fixed Effect, Random Effect, and d-Type Effect Size. In (5), the γ s are known as fixed effects since their estimated coefficients (γ s) for the individual and interaction effects are fixed. Thus, the fixed effects, γ s, are the average effects in the entire population (of schools) for corresponding individual and interaction effects, expressed in (5). Fixed effects methods completely ignore the between-teacher and between-school variation and focus only on the within-teacher and within-school variation. When the researcher wishes to investigate differences across teachers and schools in students' gain scores, it will be necessary to specify also a random effects (of teachers and schools), meaning that it is assumed that the effect varies randomly within the population and the researcher is interested to test and estimate the variance of these random effects across the population. Thus, when some effects in a statistical model are modeled as being random, we mean that we wish to draw conclusions about the population from which the observed variables were drawn, rather than about these particular variables themselves.

We analyzed both of the unconditional model and conditional VAM using PROC MIXED procedure in SAS (see [37]) (The PROC MIXED procedure is designed to fit mixed effects model where both fixed and random effects are estimated. In other words, this procedure estimates slopes of predictors and variances of random effects in multilevel model. The variances of the random-effects parameters are also known as variance components.) . This procedure estimated the fixed and random effects of model parameters. In order to answer the research question 1, the hypothesis was tested using the p values associated with the estimated fixed effects of individual predictors and cross-level interaction terms. Research question 2 was addressed by computing the d-type effect sizes using estimated variance components for teacher and school level models applying the formula provided by Rowan et al. [1] in (6)

$$d = \frac{\sqrt{(\text{Variance in gain scores lying among teachers})}}{\sqrt{(\text{Total student + teacher + school variance in student gain scores})}}. \tag{6}$$

The effect size for school level model was calculated after substituting the numerator by “variance in gain scores lying among schools” in (6).

According to Rosenthal [38], the d-type family of effect sizes are designed to express differences in outcomes across two groups (e.g., an experimental and control group) in terms of standard deviations of the outcome variable. In the context of this study, however, we analyzed data from more than two groups. In the random effects models of our analysis, the variance components are calculated from data on all of the teachers, assuming that all schools have equal variance among students and teachers. Using these variance components, we have developed a d-type effect size metric by comparing outcomes across two groups arbitrarily considering teachers with large number of students in their classes. The two comparison groups selected in our analysis are teachers within the same school who differ in their effects on student gain scores by one standard deviation. Using this procedure, the resulting d-type effect size of .22 can be interpreted as showing the difference in students’ gain scores that would be found among two students from the same school if they were assigned to teachers one standard deviation apart in effects on gain scores. For example, if the d-type effect size is .22, we would conclude that two students from the same school assigned to teachers a standard deviation apart in effectiveness would differ by .22 standard deviations in gain scores.

4. Results

Table 1 provides the significant effects of individual predictors at student, teacher, and school levels and effects of their cross-level interactions using conditional VAM for predicting students’ mathematics gain scores in order to answer the research question 1. The study found the significant effects of students’ pretest scores ($p < .0001$) and SES or socioeconomic status ($p < .0001$) on mathematics gain scores. Further, the test of hypothesis revealed that the teachers’ certification in mathematics content-area ($p = .001$) and their experience ($p = .023$) significantly predicted mathematics gain scores. The effects of continuous predictors (e.g., pretest scores, experience, and school poverty with slope estimates of .026, .042, and -4.146 , respectively) can be interpreted as the changes in mathematics gain scores associated with unit changes in respective predictors, controlling for other predictors. The slope estimates of approximately -2.0 and 2.0 were found for SES and content certification, respectively. Controlling for other predictors, the effect of SES can be interpreted as the mean mathematics gain scores of students for those who participated in free and reduced lunch (FRL) which was two times smaller than the gain scores of those students who did not participate in FRL. The effect for content certification (i.e., 2.0) can be interpreted as the mean gain scores of students taught by teachers who hold mathematics content certificates which was two times higher than the gain scores of students taught by teachers who did not hold such certificates (holding other predictors constant). Likewise, school poverty ($p < .0001$)

TABLE 1: Estimated effects of predictors for predicting mathematics gain scores in conditional model.

Effect	Estimate	Std. error	<i>t</i> value	<i>p</i> value
Intercept	19.331	0.698	27.69	< .0001
Pretest scores	0.026	0.001	26.00	< .0001
SES	-2.148	0.325	-6.61	< .0001
Content certification	1.973	0.615	3.21	.001
Experience	0.042	0.019	2.21	.023
School poverty	-4.146	0.829	-5.00	< .0001
Pretest scores * content certification	0.033	0.010	3.30	.001
SES * content certification	0.910	0.305	2.98	.003
Pretest scores * Pct. Adv. Math. Deg.	0.015	0.005	3.01	.002
SES * school mean Experience	-0.120	0.041	-2.93	.003

Note: “Pct. Adv. Math. Deg.” denotes percent of advanced mathematics degree.

had a significant impact on student mathematics gain scores with an effect of -4.146 . This can be interpreted as the factor decrease (of approximately 4) in students’ mathematics gain scores with a unit change in school percent of free and reduced lunch (i.e., school poverty).

The interaction effects of teacher’s mathematics content-area certification with students’ pretest scores ($p = .001$) and SES ($p = .003$) significantly predicted mathematics gain scores at a .05 level. The interaction effect of pretest scores and content certification was found to be .033. For example, this effect can be interpreted as the factor increase (of .033) in students’ mathematics gain scores associated with the difference in teacher’s content certification status due to one unit change in pretest scores for a reference school (i.e., for a school with average poverty). The interaction effect of content certification and SES was found to be .91. This effect can be interpreted as the factor increase (of .91) in the students’ mathematics gain scores reflecting the effect of content certification that is associated with the SES contrast.

Further, whether or not a teacher had earned an advanced degree in mathematics or in mathematics education did not significantly impact students’ gain scores. However, interaction effect between the percent of teachers in a school with advanced degrees and students’ pretest scores ($p = .002$) was found to be significant with an effect estimate of .015. A positive value for the interaction effect between these two continuous predictors implies that the higher the pretest scores, the greater effect of advanced degree (at school level) on students’ mathematics gain scores. Alternatively, the higher the percent of advanced degree (at school level), the greater the effect of students’ pretest scores on mathematics gain scores. Similarly, the interaction effect between the school mean teacher experience (i.e., average years of teacher experience at school level) and SES ($p = .003$) impacted significantly on student mathematics gain scores with an effect estimate of -0.120 . This effect can be interpreted as

TABLE 2: Estimation of variance explained, significance, and effect size at teacher and school levels.

Random effect	Variance component	Variance explained	<i>p</i> value	Effect size
<i>Teacher level effect</i>				
Unconditional model	4.500	3.6%	< .0001	0.19
Conditional VAM	4.645	4.6%	< .0001	0.22
<i>School level effect</i>				
Unconditional model	0.467	0.4%	.0410	0.06
Conditional model	0.258	0.3%	.1636	0.05

the factor decrease (of $-.120$) in students' mathematics gain scores associated with the difference in student's SES status due to one unit change in school mean (teacher) experience for a reference teacher (i.e., for a teacher with average experience, holding no content certificate and no advanced degree).

Table 2 provides the estimation of variance explained, *p* value, and d-type effect sizes for predicting mathematics gain scores in unconditional and conditional VAM at teacher level. The test of hypothesis regarding "no significant teacher-to-teacher variance in mean NCE gain scores" to predict students' mathematics gain scores, pertaining to research question 2, is rejected for both unconditional model and conditional VAM (with *p* values < .0001). For unconditional and conditional VAM at teacher level, the d-type effect sizes were .19 (3.6% variance explained) and .22 (4.6% variance explained), respectively, with an increase of .03 in effect size for the conditional VAM.

Table 2 also depicts the estimation of variance explained, *p*-value, and d-type effect sizes for unconditional and conditional models at school level. According to the results, only the unconditional model showed significant school-to-school variance (with *p* = .0410). The d-type effect sizes, at school level, were .06 (0.4% variance explained) and .05 (0.3% variance explained), respectively, for unconditional and conditional models, with a decrease of .01 in effect size for the conditional model. Thus, a trivial effect size was found at school level for both unconditional and conditional models. Given the above results for conditional model at school level that the variance explained was not significant and a trivial amount of effect size was found, it is shown that the school level factors were relatively less important for measuring teacher effectiveness.

5. Discussion

The findings of this study have several implications. Discussing about student level predictors in the model, the findings showed positive impact of students' prior status scores and negative impact of their socioeconomic status on students' gain scores. It is not surprising that high achieving students in prior year will tend to have high gain scores in current year. However, since the purpose of this study is to predict student gain scores using value-added model (which considers the adjustment of student level covariates such as prior scores and background variables) at student level, we

have examined the effects of these predictors in the model. The magnitude of effect sizes at teacher level (models) using these predictors are similar to the effect sizes reported by Rowan et al. [1].

Several teacher level factors were important for determining teacher effectiveness. For example, teacher content-area certification had significant positive impact on students' gain scores. This implies that the schools should focus on hiring the teachers who have content-area certification in mathematics in order to increase students' gain scores. This finding also concurs with the results from past research (see [17]). Further, significant positive effects on gain scores were found due to interactions of content-area certification with students' pretest scores and SES. Thus, we can claim that the teachers holding content-area certification will be able to increase student achievement gains after the interaction with student level predictors. Another piece of evidence that the teachers with content-area certification have key role in increasing students' gain scores is that this factor produced significant positive interaction effect after interacting with SES. Although SES showed significant negative effect originally (in level-1 model), it has been changed from negative to positive (effect) after the interaction with content-area certification, which is an important implication. Further, it can be claimed that more experienced teachers are instrumental in increasing students' gain scores according to the findings of this study.

At school level, school poverty showed significant negative effect on students' gain scores. However, significant positive effect is found due to the interaction of percent of advanced degree in mathematics or mathematics education with pretest scores. This implies that given students' prior achievement, the greater the percent of advanced degree teachers in mathematics-related field in a school is, the higher would be students' gain scores. School teacher mean experience showed a significant negative effect on student achievement gain while interacted with SES. This means that given the SES of students, the schools with concentrations of teachers with rich experience did not help increase students' gain scores.

Both unconditional model and conditional VAM at teacher level showed significant variances and moderate effect sizes. However, comparing the effect sizes of both models (in Table 2), the conditional VAM with predictors at teacher level model is preferable over the unconditional model since the former model produced larger effect size

than the later model. At school level, both unconditional and conditional models explained small percents of variance and, consequently, the effect sizes of negligible magnitude were produced. To our surprise, the conditional model also showed a trivial effect size at school level even after including the factors such as school poverty and other significant interaction effects at this level. Since a trivial effect size of .05 was produced due to the conditional model at school level, the factors at school level have a very little contribution to teacher effectiveness. This implies that during the process of measuring teacher effectiveness, we should not take the percent of free and reduced lunch at school level into consideration with significant emphasis.

The other important finding of this research is that if we assign the teachers with mathematics content-area certification to teach impoverished students in a school, then these teachers can increase students' gain scores. Teaching impoverished students in a school by teachers with many years' of experience in teaching, however, did not help to produce effective results in terms of increasing students' gain scores.

It is relevant here to discuss the recent research work on value-added model by Rothstein [39] who developed falsification tests based on the assumption that teachers in later grades (e.g., fifth grade) cannot have causal effects on students' test scores in earlier grades (e.g., fourth grade). He also mentions that students' gains are affected by school as well as nonschool factors and allowing non-random assignment of students (to classrooms/teachers) limits the interpretation of teacher effect on student gain scores. Acknowledging the limitation of non-random assignment of students to teachers, we also recognize that this study is limited to incorporating selected school-related factors at student, teacher, and school levels for predicting students' gain scores.

6. Conclusions

This study employed a three-level HLM using unconditional model and conditional VAM to predict mathematics gain scores in middle schools. Such models were employed in order to measure the effects of student, teacher, and school level predictors and examine the magnitude of d-type effect sizes at teacher and school level models.

The findings indicated significant positive effects of middle school teachers' mathematics content-area certification, teacher experience, and the interaction effects of content-area certification with students' pretest scores and SES. The findings of this study imply that the teacher quality, represented by teacher content-area certification in mathematics and teacher experience as well as interaction effects associated with these predictors, is an important factor in predicting mathematics gain scores in middle schools. We found that the conditional VAM produced larger effect size than that of unconditional model at teacher level. Further, the effect sizes associated with school level model were trivial although school poverty and some other interaction effects (associated with school poverty)

showed significant impact on students' mathematics gains scores. This study provided the evidence that school level factors are relatively less important for measuring teacher effectiveness.

This research provides important information on teacher and school evaluations for schools, school districts, and the Department of Education in the states. First, given the significant effects of relevant predictors to measure teacher effectiveness, the results will be beneficial as the potential predictors can be controlled in order to increase gain scores and reform schools. Second, evaluators and researchers can replicate similar conditional VAM in order to measure teacher effectiveness in their context.

7. Limitations and Recommendations for Further Research

The authors recognize that there are several limitations associated with this study. First, students and teachers are not randomly assigned to classrooms and there is no statistical model which can fully make up for this lack of randomization. Another limitation to this research is that seniority may not be a true measure of experience because seniority is a measure of how many years a teacher has been employed in the school district in current study. If there are errors associated with this limitation, the seniority (expressed in number of years) will underestimate the number of true years of experience. However, since it is a common trend to assume the teacher experience as the number of years the teacher involved in teaching, we should be comfortable with valid interpretation of the results associated with this predictor's effect. Further, there may be a masked effect of advanced degree in mathematics while measuring the effect of content-area certification because most of the teachers with such degrees may have content-area certification in pertinent subject. We also know that there are certainly more descriptors of student SES than just free and reduced lunch participation. However, provided the specific interpretation related to the effect of SES, the study results can be validly generalized in relevant contexts. It should be noted that seniority and free and reduced lunch variables were the only determinants available across all schools involved in the study. Finally, students' mathematics gain scores on the FCAT were the only outcome measure used for student achievement since it is the only state mandated standardized test across all districts in the state of Florida.

Future studies are suggested to cover more grades and more school districts since this research is limited only within middle schools in a large urban school district. Researchers are also recommended to describe the teacher effectiveness based on simple effects, as demonstrated by Subedi [34], in addition to d-type effect size.

Acknowledgments

The earlier version of this paper was presented in the annual meeting of the American Educational Research Association (AERA), April 30–May 4, 2010, Denver, Colorado, U.S.A. We

acknowledge the anonymous ERI reviewers for their valuable comments.

References

- [1] B. Rowan, R. Correnti, and R. J. Miller, "What large-scale, survey research tells us about teacher effects on student achievement: insights from the prospects study of elementary schools," *Teachers College Record*, vol. 104, no. 8, pp. 1525–1567, 2002.
- [2] P. F. Tobe, "Value-added models of teacher effects," in *International Handbook of Research on Teachers and Teaching*, L. J. Saha and A. G. Dworkin, Eds., pp. 1113–1134, Springer, New York, NY, USA, 2009.
- [3] H. R. Jordan, R. Mendro, and D. Weerasinghe, "Teacher effects on longitudinal student achievement: a preliminary report on research on teacher effectiveness," Paper presented at the National Evaluation Institute, Indianapolis, Ind, USA, 1997.
- [4] W. Sanders and J. Rivers, "Cumulative and residual effects of teachers on future student academic achievement," Tech. Rep., University of Tennessee, Value-Added Research and Assessment Center, Knoxville, Tenn, USA, 1996.
- [5] J. Scheerens and R. Bosker, *The Foundations of Educational Effectiveness*, Pergamon, New York, NY, USA, 1997.
- [6] M. B. Allen, "Eight questions on teacher recruitment: what does the research say?" Denver: Education Commission of the States, 2005, <http://www.ecs.org/html/educationissues/teachingquality/trrreport/home/TeacherRecruitmentRetention.pdf>.
- [7] J. R. Betts, K. S. Rueben, and A. Danenberg, *Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California*, Public Policy Institute of California, San Francisco, Calif, USA, 2000.
- [8] E. A. Hanushek, J. F. Kain, and S. G. Rivkin, "Why public schools lose teachers," *Journal of Human Resources*, vol. 39, no. 2, pp. 326–354, 2004.
- [9] H. Lankford, S. Loeb, and J. Wyckoff, "Teacher sorting and the plight of urban schools: a descriptive analysis," *Educational Evaluation and Policy Analysis*, vol. 24, no. 1, pp. 37–62, 2002.
- [10] U.S. Department of Education (USDOE), Office of Postsecondary Education, "The Secretary's fourth annual report of teacher quality: a highly qualified teacher in every classroom," Tech. Rep., USDOE, Washington, DC, USA, 2005.
- [11] C. D. Jerald and R. M. Ingersoll, "All talk, no action: putting an end to out-of-field teaching," Tech. Rep., The Education Trust, Washington, DC, USA, 2002.
- [12] K. Haycock, "Good teaching matters: how well-qualified teachers can close the gap," Thinking K-16, 3, 1–14, The Education Trust, Washington, DC, USA, 1998.
- [13] E. Greenberg, D. Rhodes, X. Ye, and F. Stancavage, "Prepared to teach: teacher preparation and student achievement in eighth-grade mathematics," in *Proceedings of the Annual Meeting of the American Educational Research Association*, San Diego, Calif, USA, 2004.
- [14] J. K. Rice, *Teacher Quality: Understanding the Effectiveness of Teacher Attributes*, Spencer Foundation, Chicago, Ill, USA, 2003.
- [15] H. Wenglinsky, "How schools matter: the link between teacher classroom practices and student academic performance," *Educational Policy Analysis Archives*, vol. 10, no. 12, 2002.
- [16] R. G. Croninger, J. K. Rice, A. Rathbun, and M. Nishio, "Teacher qualifications and early learning: effects of certification, degree, and experience on first-grade student achievement," *Economics of Education Review*, vol. 26, no. 3, pp. 312–324, 2007.
- [17] L. Darling-Hammond, D. J. Holtzman, S. J. Gatlin, and J. V. Heilig, "Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness," *Education Policy Analysis Archives*, vol. 31, no. 42, 2005.
- [18] L. Darling-Hammond, "Teacher quality and student achievement: a review of state policy evidence," *Education Policy Analysis Archives*, vol. 8, no. 1, 2000.
- [19] P. T. Decker, D. P. Mayer, and S. Glazerman, *The Effects of Teach For America on Students: Findings from a National Evaluation*, Mathematica Policy Research, Princeton, NJ, USA, 2004.
- [20] T. J. Kane, J. E. Rockoff, and D. O. Staiger, "What does certification tell us about teacher effectiveness? Evidence from New York City," Tech. Rep., National Bureau of Economic Research, Cambridge, Mass, USA, 2006.
- [21] D. D. Goldhaber and D. J. Brewer, "Does teacher certification matter? High school teacher certification status and student achievement," *Educational Evaluation and Policy Analysis*, vol. 22, no. 2, pp. 129–145, 2000.
- [22] D. Ballou and M. Podgursky, "Reforming teacher preparation and licensing: what is the evidence?" *Teachers College Record*, vol. 102, no. 1, pp. 5–27, 2000.
- [23] D. D. Goldhaber and D. J. Brewer, "When should we reward degrees for teachers?" *Phi Delta Kappan*, vol. 80, no. 2, pp. 134–138, 1998.
- [24] C. Howley, "Compounding disadvantage: the effects of school and district size on student achievement in West Virginia," *Journal of Research in Rural Education*, vol. 12, pp. 25–32, 1996.
- [25] L. Lippman, S. Burns, and E. McArthur, "Urban schools: the challenge of location and poverty," Tech. Rep., Department of Education, Office of Educational Research and Improvement, Washington, DC, USA, 1996.
- [26] D. H. Monk, "Subject area preparation of secondary mathematics and science teachers and student achievement," *Economics of Education Review*, vol. 13, no. 2, pp. 125–145, 1994.
- [27] B. Swan, *Middle school mathematics teacher certification, degree level, and experience, and the effects on teacher attrition and student mathematics achievement in a large urban district*, Doctoral dissertation, University of Central Florida, Orlando, Fla, USA, 2006, publication no. AAT 3233679.
- [28] B. Swan and B. R. Subedi, "Middle-school teacher preparation, certification and students' mathematics achievement in a large urban district," Tech. Rep., The Multi-University Reading, Mathematics and Science Initiative. Learning System Institute, Florida State University, 2005.
- [29] B. R. Subedi, "Predicting reading proficiency in multilevel models: an ANOVA-like approach of interpreting effects," *Educational Research and Evaluation*, vol. 13, no. 4, pp. 327–348, 2007.
- [30] K. Banks, "The effect of school poverty concentration in WCPSS," Evaluation and Research Department, Wake County Public School System, 2001, www.wcpss.net/evaluation-research/reports/2001/0123_Poverty.pdf.
- [31] B. R. Subedi and B. Swan, "Predicting student gains with hierarchical models: a value-added approach to measure teacher effectiveness," in *Proceedings of the Annual Meeting of American Educational Research Association*, San Diego, Calif, USA, April 2009.
- [32] H. Goldstein, "State face challenges measuring academic growth that education's initiatives may help address. Report to

- Congressional requesters United States,” in *Multilevel Statistical Models*, John Wiley & Sons. Government Accountability Office (GAO), New York, NY, USA, 2nd edition, 1995.
- [33] S. W. Raudenbush and A. Bryk, *Hierarchical Linear Models in Social and Behavioral Research: Applications and Data Analysis Methods*, Sage, Newbury Park, Calif, USA, 2nd edition, 2002.
- [34] B. R. Subedi, *A Demonstration of three-level hierarchical generalized linear model applied to educational research*, Doctoral dissertation, Florida State University, Tallahassee, Fla, USA, 2005.
- [35] C. S. Berkey, D. C. Hoaglin, F. Mosteller, and G. A. Colditz, “A random-effects regression model for meta-analysis,” *Statistics in Medicine*, vol. 14, no. 4, pp. 395–411, 1995.
- [36] C. Morris and S. Normand, “Hierarchical models for combining information and for meta-analysis,” *Bayesian Statistics*, vol. 4, pp. 321–344, 1992.
- [37] J. D. Singer, “Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models,” *Journal of Educational and Behavioral Statistics*, vol. 23, no. 4, pp. 323–355, 1998.
- [38] R. Rosenthal, “Parametric measures of effect size,” in *The Handbook of Research Synthesis*, H. Cooper and L. V. Hedges, Eds., Russell Sage Foundation, New York, NY, USA, 1994.
- [39] J. Rothstein, “Do value-added models add value? Tracking, fixed effects, and causal inference,” Tech. Rep. 159, Princeton University, Princeton, NJ, USA, 2007, <http://www.princeton.edu/~ceps/workingpapers/159rothstein.pdf>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

