Value Added Assessment of Teacher Preparation in Louisiana: 2005-2006 to 2007-2008

George H. Noell, Ph.D.
Department of Psychology

Kristin A Gansle, Ph.D.
Department of Educational Theory, Policy, & Practice

R. Maria Patt, M.A.
Michael J. Schafer, B.A.
Department of Psychology

Louisiana State University

August 11, 2009

Acknowledgements

This report is based upon data provided by the Division of Planning, Analysis, and Information Resources of the Louisiana Department of Education. The authors would like to thank Michael Collier, Allen Schulenberg, and David Elder (Division Director), without whom this work would not have been possible. Additional data were provided by the Louisiana Board of Regents.

Any errors, omissions, or misstatements contained herein are entirely the responsibility of the authors. Any conclusions proffered are the responsibility of the authors and do not reflect the views of the Louisiana Department of Education or the professionals from that organization who provided professional guidance and technical assistance.

This work was supported by award CT-06/07-VAA-01 from the Louisiana Board of Regents, which was funded by a grant from the Carnegie Corporation of New York.

Table of Contents

Abstract

Value Added Assessment of Teacher Preparation

Analyses were conducted examining the degree to which the educational attainment of students taught by recent graduates of specific teacher preparation programs either met, failed to meet, or exceeded expectations based on prior achievement and demographic factors as compared to experienced teachers. Work began with the construction of a large multivariate longitudinal database linking data about students, teachers, and courses over four academic years. This was followed by a model development phase in which hierarchical linear models were developed to predict student achievement based upon prior achievement, student demographic factors, and classroom level covariates. The models nested students within teachers and teachers within schools. Separate models were developed for each content area and school year. These models were used to assess the efficacy of teacher preparation programs. Analyses were conducted across a pooled data set spanning the academic years 2005-2006, 2006-2007, and 2007-2008. Due to the timing of teacher preparation program (TPP) redesign and the meaning of the data relevant to current programs, results are limited to redesigned teacher preparation programs that had a sufficient number of graduates teaching in assessed subjects and grades. As a result, all of Louisiana's teacher preparation programs are not yet represented in the report. As redesigned programs continue to operate and produce new graduates, the number of programs represented in subsequent reports will increase substantially. Effect estimates were placed into five performance bands that were developed to describe teacher preparation programs. With a few exceptions, results from this year's assessment are generally consistent with the 2008 report. TPPs generally fell within the same performance bands across this report and the 2008 report within individual content areas. Thirty-nine TPP by content area combinations had sufficient data to be included in this report; last year, 30 combinations had sufficient data.

Technical Report:
Value Added Assessment of Teacher Preparation in Louisiana:
2005-2006 to 2007-2008

I. Introduction

This report describes the results of the *Value Added Assessment of Teacher Preparation Project* (VAA-TPP) for the academic years 2005-2006, 2006-2007, and 2007-2008. These analyses build upon results reported previously in Noell (2006), Noell, Porter, and Patt (2007), and Noell, Porter, Patt, and Dahir (2008). The VAA-TPP project is a program evaluation study housed in the Department of Psychology at Louisiana State University. The VAA-TPP is building longitudinal databases linking students across years and linking those students to their teachers in core content areas. These longitudinal databases are then used to assess the impact of teacher preparation programs (TPP) on the educational attainment of students taught by their graduates.

At this stage in its development, the VAA-TPP examines the average impact of new teachers from specific preparation programs. The research team does not have data sufficient to examine the differential effects of TPP in domains such as recruitment, admissions, content preparation, pedagogical preparation, field experiences, screening for graduation, or transition into the workforce. Additionally, the assessment examines the mean effect for graduates from these programs in specific content areas. It does not provide data regarding the efficacy of individual teachers. A separate statewide research team led by Dr. Jeanne Burns that includes representatives from all TPPs in Louisiana is currently collecting data examining the process of teacher preparation. These data are being integrated into the longitudinal data that are the basis of this report, and will provide the foundation for efforts to examine the process of teacher preparation within the VAA-TPP. Results of these process analyses will be provided separately from this document.

In the context of this report, *value added analysis* (VAA) describes the use of demographic and prior achievement data to estimate expected outcomes for students in a specific content domain (e.g., Mathematics) based on a longitudinal data set derived from all students who took state mandated tests in grades 3 through 9 in Louisiana. The assessment uses a relatively complex model that includes the grouping of students within classrooms and classrooms within schools. The model then examines the degree to which students who are taught by new teachers from specific TPPs compare to other students after controlling for prior achievement and demographic factors. This information is used to estimate the degree to which new teachers' effectiveness is differentially associated with having entered teaching through specific TPPs.

The estimation of educational effects within complex longitudinal models that can accommodate the correlation of errors that emerge due to the nesting of students within classrooms is a literature base that is beyond the scope of this report (for example, see: Ballou, Sanders, & Wright, 2004; Goldhaber & Brewer, 1997; Hill, Rowan, & Lowenberg, 2005; Hong & Raudenbush, 2008; McCaffrey, Lockwood, Kortez, &

Hamilton, 2003; McCaffrey, Lockwood, Kortez, Louis, & Hamilton, 2004; Todd & Wolpin, 2003; Wayne & Youngs, 2003). This technical report summarizes the findings of the analyses through 2007-2008.

Prior Work

The methods employed in this report were derived in prior research (Noell, 2005; Noell, 2006; Noell et al., 2007; Noell et al., 2008). Data relevant to analytic decisions and rationale are provided in those reports. The assessment model is based on hierarchical linear models (HLM; McCulloch & Searle, 2001; Raudenbush & Bryk, 2002) that nested students within teachers and teachers within schools, and as a result permit correlation of error terms within nested units. This allows for modeling of variables at the student, teacher, and school level in a methodologically appropriate manner. The nesting structure also permits specifying a model in which effects such as those of schools upon teachers who in turn affect students can be appropriately linked through the hierarchy.

The prior work examined a number of specific issues in the specification of the assessment models. For example, based on examination of estimated teacher effects by years of experience cohorts, new teachers were defined as first *and* second year teachers (Noell et al., 2007). Additionally, the minimum standard for reporting results for an individual university programs was set at 25 observations per program of teacher/year outcomes. This is based on an examination of the ratio of variance within program estimates to the variance between programs relative to the number of graduates (see Noell et al., 2007, for a detailed discussion).

One of the most important modeling conventions adopted within the prior work was the decision to use a single year covariate adjustment approach for modeling student achievement (Noell, 2006; Noell et al., 2007). This approach uses five achievement test scores from the prior year combined with more than 12 demographic variables to predict current year achievement. Although these models have extensive specifications that account for a substantial portion of the variance in student achievement, they do not capitalize on the analytic power and elegance of multiyear achievement trajectories for students across multiple teachers (see McCaffrey et al., 2003; McCaffrey et al., 2004; Nye, Konstantopoulos, & Hedges, 2004; Sanders & Horn, 1998; Todd & Wolpin, 2003).

The decision to use a covariate adjustment approach was guided by two considerations. First, the covariate adjustment models were able to account for a substantial and credible portion of the variance in achievement, suggesting that they are sufficient for this type of assessment. Second, multiyear, repeated observation models generally assume that the quantity that is being observed across years is an unchanging one-dimensional scale such as dollars or truly vertically aligned educational tests (Matrineau, 2006; Seltzer et al., 1994). Although there can be considerable debate about the degree to which vertical scaling is actually achieved or is achievable in educational assessment over wide grade spans (see Matrineau et al., 2007; Reckase, 2004), a plausible argument cannot be made that Louisiana's assessments are vertically aligned. The tests are aligned to the content standards for each grade and as a result are an assessment of the blueprint of instruction. However, that means that the specific content and weighting of the content represented on the instruments shifts considerably from one year to the next. This is particularly striking in Science and Social Studies where some years are thematically focused (e.g., Life Science or Louisiana history). Interested readers can examine http://www.doe.state.la.us/lde/saa/2273.html for a description of Louisiana's assessment content by grade level. A covariate adjustment model can be built upon relatively modest assumptions regarding the measurement properties of the tests that contribute to them (see Martineau et al., 2007; Reckase, 2004; Seltzer, Frank, & Bryk, 1994), and these assumptions are tenable for Louisiana's tests.

The treatment of students who are retained is another substantial benefit of this approach. A single year covariate model does not accentuate the lost records/linkages problems that arise from grade retention (which is a significant issue in Louisiana due to high rates of retention). Obviously, a student taking the 4th grade assessment in two consecutive years cannot be analyzed jointly with students who are taking tests at two different grade levels. The analyses reported here replicate prior work in which an HLM covariate approach to the data (Noell, 2006; Noell et al., 2007; Noell et al., 2008) was adopted.

Programs Included in the Current Report

In order for a program to be included in the assessment it has to complete the training of new teachers (1 to 5 years typically), and a sufficient number of those teachers have to complete at least one year of teaching in a public school in a tested subject and grade. The minimum number of teachers who are represented in the data set required for inclusion is 25, based on prior research (Noell et al., 2007). For most programs, this will require considerably more than 25 graduates due to a variety of factors.

For purposes of illustration, assume that a TPP had 100 graduates in a particular year. Of these graduates, some will teach subjects such as band, foreign language, or physical education. Assuming that 20% of the graduates teach in these areas, 80 new teachers would remain whose effects on student achievement theoretically could be estimated. Of the 80 new teachers, some will not enter public school teaching. They will teach in private schools, pursue graduate study, delay work entry to start families, or pursue employment outside schools. This part of the attrition could readily reduce the number of available new teachers to 50. Of this number, half will typically teach outside tested grades and half will teach in tested grades. Of this 25, assume approximately 13 teachers teach all subjects in the elementary grades and 12 teach a single content in middle school or high school (i.e., 3 teachers per content area). If this pattern held, there would be 16 teachers per content area in each year's cohort. The assessment model capitalizes observations of teachers across three years, so in this assessment, two graduate cohorts would be required for the TPP to be included in the analysis.

Due to the redesign of Louisiana's TPPs during the period 2000 to 2003, many of the new teachers who have entered the workforce completed programs that have since been retired and are not the focus of this assessment. However, with the 2007-2008 school year, the first large scale entry of post-redesign undergraduate program completers was evident in the work force. For most programs, there were not enough graduates in a single year to be included in the assessment. However, a number of them had enough new teachers in 2007-2008 that it appears likely that they will be included in the next report. That report will be released in 2010.

II. Data Merging Process

Data for the academic years described in previous reports were merged following a process that was substantially replicated with the current year data (see Noell, 2006; Noell et al., 2007; Noell et al., 2008). The data from individual school years were then combined to form a larger multiyear data set (described below) for the purpose of assessing TPPs.

Data for 2007-2008 were drawn from the standardized test files (*i*LEAP and *LEAP-21*) for spring 2007 and 2008, the Louisiana Educational Accountability Data System (LEADS, formerly Curriculum database) linking students to teachers, and supplemental student databases. The testing and supplemental databases provided data regarding attendance, enrollment, disability status, free lunch status, and demographic variables (e.g., race and gender). Data regarding teachers were drawn from the certification database, teacher attendance, and teacher demographic databases obtained from the Louisiana Department of Education. Additionally, all TPP completers were identified through data provided to the Board of Regents by the TPPs. A multistage process was used to create longitudinal records for students describing achievement, attendance, and demographic factors across years. Similarly, teacher data were merged to create complete records for preparation, attendance, and certification. The student and teacher databases were then linked through LEADS.

Initial work was undertaken to resolve duplicate records and multiple partially complete records that described the same student within separate databases. Following this work, data files were merged in a series of steps and a further round of duplication resolution was undertaken. Students' data were linked across years based upon unique matches on multiple identifiers used in each stage of the matching process. Student records that remained unmatched were then examined for a potential unique match through a layered series of comparisons. The matching process included six stages that were implemented hierarchically and that required unique matches on at least three identifying variables in order for a match to be established. Additional details of this process are available from the first author.

Table 1 describes the number of records available and the percentage of the total records that were matched at that stage. Mathematics and Science are provided as examples of the merging process as English-Language Arts (ELA) is similar to Mathematics and Social Studies is similar to Science. The difference between these clusters is the result of an assessment in 9th grade in Mathematics and ELA, but not in Science and Social Studies.

Several important decision points are noteworthy. Initial records were limited to students who completed one assessment in grades 4-9 to permit the availability of one year prior achievement data. The testing program begins in the 3rd grade, so 4th graders would have their matched 3rd grade achievement data as predictors of 4th grade achievement. Although the proportion of matches between the years is large, there is some attenuation due to several factors. In order to be included in the analyses, a student was required to be enrolled in the same school from September 15, 2007 to March 15, 2008. Because the student-teacher-course nexus data are collected only once per year, once a student changes schools within that time period, it is not possible to ascribe achievement measured at the end of that period to a particular teacher. The records available for analysis were further attenuated by the number of students whose matched data were not from consecutive grades (e.g., 3rd to 4th). Some students were retained in grade or promoted two grades in a single year. Obviously, the meaning of taking the same test two years in a row or completing assessments separated by more than one grade level differs from taking tests in the expected sequence. As a result they were excluded from analyses. Finally, in order to be included in the analyses, the students' attendance and achievement records had to be matched to the LEADS curriculum data to identify which courses the students took and who taught those courses. Additionally, the attendance and course databases had to confirm that the student was enrolled in the same site.

Table 1: *Cases Available at Each Stage of the Matching Process*

	Mathematics	**Science**
Assessed students grades 4-9 in 2008	295,810	248,175
Matched to 2007 data	272,134 (92.0%)	231,069 (93.1%)
Consecutive grades assessed	250,188 (84.6%)	212,644 (85.7%)
Single primary school of attendance In curriculum database	233,336 (78.9%)	198,085 (79.8%)

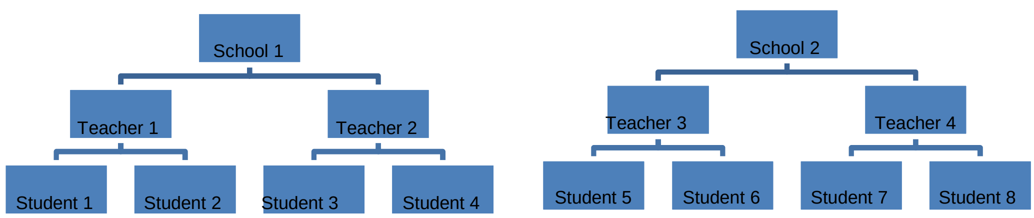
Table note. The percentage in parentheses within each cell is the percentage of the total records available for analysis in that content area at that stage of database construction.

Once students' achievement, demographic, attendance, and course enrollment records were linked, these data were linked to information about their teachers. This included teacher certification data obtained from the Louisiana Department of Education's Division of Planning, Analysis, and Information Resources and TPP data obtained from the Louisiana Board of Regents. Course codes were collapsed into groups that were associated with the specific test areas (i.e., Mathematics, Reading, English-Language Arts, Science, and Social Studies). For example, 4th grade Reading was associated with Reading test scores and Life Science with Science test scores. Course codes that could not reasonably be linked to a standardized test (e.g., Jazz Ensemble) were dropped. Students who had more than one teacher in a content area were included for each teacher, but their weight was reduced in proportion to the number of classes in that content area in which the student was enrolled. For example, if a student was enrolled in two Mathematics classes, that student would have a record linked to each Mathematics teacher, but each was weighted 0.5, or contributed ½ of the amount that a student with only one class contributed to a single teacher. This convention was also used to account for team teaching.

III. Building the Base Model of Student Achievement Prior to VAA

Replicating the approach used in Noell (2006), Noell et al. (2007), and Noell et al. (2008), the educational assessment data were analyzed using hierarchical linear models (HLM; McCulloch & Searle, 2001; Raudenbush & Bryk, 2002). Hierarchical models were developed with students nested within teachers that were in turn nested within schools. Interested readers may choose to consult Noell et al. (2007) for a detailed discussion of the variance apportionment between levels of the model, alternative models, and the impact of using a covariate adjustment approach to modeling results. This information will not be repeated here. Figure 1 below depicts the nesting structure that was employed.

Figure 1: *Nesting Structure of Students with Teachers and Teachers within Schools*



Building the current models. The general strategy of the modeling approach used was somewhat parallel to Tekwe and colleagues (2004) and previously has been followed by the VAA-TPP. Model development was completed independently for each school year: 2005-2006, 2006-2007, and 2007-2008. Please consult previous reports for descriptions of model development during those years. This report describes results of model development for the 2008 assessments. The approach was replicated across

Mathematics, Reading, English-Language Arts, Science, and Social Studies. Error at each of the three levels (student, teacher, and school) was assumed to be normally distributed with a mean of 0 and common variance at that level. An initial 3-level model was specified in which achievement was modeled with no prior predictors as a basis for comparison with more complex models. Students' prior achievement in English-Language Arts, Mathematics, Reading, Science, and Social Studies were entered in a block as fixed effects. All effects were significant in all content areas and were retained. Next, the 16 demographic variables presented in the table below were entered as a block. Variables were then removed one at a time in order of the lowest *t* value until all remaining effects were significant at $p < .01$. Variables examined are listed in Table 2.

Table 2: *Student Level Variables*

Variable
Gender (Male)
African American
Asian American
Native American
Hispanic
Emotionally Disturbed
Speech and Language
Mild Mental Retardation
Specific Learning Disability
Other Health Impaired
Special Education - Other
Gifted
Section 504
Free Lunch
Reduced Price Lunch
Student Absences
Prior Year Mathematics Test
Prior Year Reading Test
Prior Year Science Test
Prior Year Social Studies Test
Prior Year English English-Language Arts Test

The decision to include student absences in the model will be evaluated as problematic by some readers. Some teachers will influence the level of student absences by the manner in which they teach and interact with students. This can result in higher or lower levels of absence. However, given that the students contributing to the analyses are minors typically between 8 and 15 years of age, their choice in whether or not to attend school is generally strongly bounded by parental intervention. This is not so much an issue of absolute contribution but of relative contribution to student absence. The authors adopted the assumption that students' absences are likely to be determined to a greater extent by variables that are beyond teacher control such as illness, parental choice, and chronic truancy than they are by student-teacher interaction. As a result, student absences were retained as a potential predictor of student achievement.

Once a model for student level achievement was developed, several classroom/teacher variables were examined. These variables were entered at the classroom/teacher level and were conceptualized as contextual factors that may moderate student achievement in addition to teachers. The classroom/teacher variables that were examined are presented in Table 3.

Table 3: *Classroom/Teacher Level Variables*

Variable
Percentage of students who were male
Percentage of students who were minorities
Percentage of students who received free lunch
Percentage of students who received reduced price lunch
Percentage of students who were in special education
Percentage of students who were identified as gifted
Percentage of students who exhibited limited English proficiency
Percentage of students identified as protected by Section 504
Class mean prior achievement in English-Language Arts
Class mean prior achievement in Reading
Class mean prior achievement in Mathematics
Class mean prior achievement in Science
Class mean prior achievement in Social Studies
Teacher absences

As with the student level demographic factors, these classroom variables were entered in a block and removed one at a time in order of smallest *t* value for the coefficient. Once all effects were significant at the .01 level, the model for that content area was finalized. The same modeling process was then implemented across content areas for level 3 of the model (schools). The variables that were initially entered in a block are listed in Table 4.

Table 4: *School Level Variables*

Variable
Percentage of students who were male
Percentage of students who were minorities
Percentage of students who received free lunch
Percentage of students who received reduced price lunch
Percentage of students who were in special education
Percentage of students who were identified as gifted
Percentage of students who exhibited Limited English Proficiency
Percentage of students identified as protected by Section 504
School mean prior achievement in English-Language Arts
School mean prior achievement in Reading
School mean prior achievement in Mathematics
School mean prior achievement in Science
School mean prior achievement in Social Studies

Tables 5 through 9 present the variables that were retained at the student, teacher, and school levels for each content area prior to consideration of teacher preparation effects. In all cases, models were developed for intercepts as outcomes. At level 1 (students), prior achievement, demographic variables, and attendance were retained as predictors of test performance. At level 2, (teachers) classroom covariates were entered as predictors of the level 1 intercept (classroom mean) only and this effect was modeled as random. No classroom level predictors were entered for student level coefficients and student level coefficients were fixed. At level 3 (schools), school building level covariates were entered as predictors of the classroom intercept (school mean) only and this effect was modeled as random. No school building level predictors were entered for classroom level coefficients, and classroom level coefficients were fixed. These model specifications were adopted to enhance the interpretability of the data and were guided by the current research questions.

In summary, classroom and school building level covariates were used to adjust intercepts for students and classrooms respectively. No covariates were used to predict lower level coefficients and all coefficients were treated as fixed. Error variance was modeled for intercepts only. A simplified presentation of the model is provided below. Only equations for intercepts are presented. All other equations (e.g., the level 2 and level 3 models for level one coefficients) were modeled as fixed and not varying. In the equations presented below, $\sum$ is used to indicate summing across the p, q, and s coefficients at the student, teacher, and school levels of the model respectively.

Level 1: Students

$$Y_{ijk} = \pi_{0jk} + \sum(\pi_{pjk})a_{pijk} + e_{ijk}$$

where

Y_{ijk} is the achievement of student i in class j at school k in the target subject

π_{0jk} is the mean achievement for classroom j at school k

π_{pjk} are the *p* coefficients that weight the contribution of the student level data in the prediction of Y for *p* = 1 to the total number of coefficients

a_{pijk} are the student level data (prior achievement, demographic variables, and attendance) that predict achievement for *p* = 1 to the total number of data points

e_{ijk} the student level random effect, the deviation of the predicted score of student i in classroom j in school k from the obtained score

Level 2: Classrooms

$$\pi_{0jk} = \beta_{00k} + \sum(\beta_{q0k})X_{q0jk} + r_{0jk}$$

where

π_{0jk} is the mean achievement for classroom j at school k

β_{00k} is the mean achievement for school k

β_{q0k} are the q coefficients that weight the weight the relationship between the classroom characteristics and π_{0jk}, q = 1 to the total number of coefficients

X_{q0jk} are the classroom level data that are used to predict achievement; this is also the location in the model at which codes for recent TPP completers are entered (described below)

r_{0jk} the classroom level random effect, the deviation of classroom jk's measured classroom mean from its predicted mean

Level 3: Schools

$$\beta_{00k} = \gamma_{000} + \sum(\gamma_{s00})W_{s00k} + u_{00k}$$

where

β_{00k} is the mean achievement for school k

γ_{000} is the grand mean achievement in the target subject

γ_{s00} are the s coefficients that weight the weight the relationship between the school characteristics and β_{00k} for s = 1 to the total number of coefficients

W_{s00k} are the school level data that are used to predict achievement

u_{00k} the school level random effect, the deviation of school k's measured classroom mean from its predicted mean

The values presented in the tables below are the final values that were obtained prior to entering teacher preparation program codes into the model. The coefficients for university preparation programs are presented in the section regarding the VAA of teacher preparation.

The coefficients are scaled to the approximate standard deviation of the educational assessments (*i*LEAP and *LEAP*) used in Louisiana: 50. As in previous years, generally, the previous year's achievement for a student in a given content was the largest predictor of the current year's achievement in that content area. It is important to note that differences in how variables were scaled create the need for considerable caution in comparing the coefficients across different types of predictors. Demographic variables at the student level were coded 1 if present and 0 if absent. Prior achievement is measured in standard deviation units from the grand mean prior achievement. Classroom percentages are measured in 10% units, so that the value presented would be the expected change in students' scores if the percentage of the indicated group increased by ***10%***. Due to differences in scales of measurement and the meaning of the measurements it is difficult to make direct comparisons across different types of measures.

It is important to recognize that the inclusion of teacher absences in the model will be regarded as problematic by some readers. To the extent that TPPs are more or less successful in preparing teachers who have poor or excellent work attendance, this variable could be siphoning off some of the TPP effect. However, it may also be the case that factors beyond the control of TPPs are likely to be more determinative regarding teacher attendance. In particular, teacher health and school district professional development requirements seem likely to have a larger impact on attendance than TPPs.

Although the coefficients at the classroom and building levels are often large, they do not lend themselves to easy interpretation. These classroom and building level variables account for residual variance in models that are saturated with student level variables that have absorbed the vast majority of the variance in the current year's achievement. As a result, the direction of the effect is counterintuitive in some cases. This finding may not be surprising when one considers the small amount of variance accounted for by the school building level of the model and the tremendous amount of information provided at the student level.

Table 5: *Hierarchical Linear Model for Mathematics Achievement*

Model Level	Variables Entered	Coefficient	(CI)
Student level variables	Gender (Male)	2.2	(1.9, 2.4)
	African American	-4.8	(-5.2, -4.5)
	Asian American	5.7	(4.7, 6.8)
	Native American	-2.2	(-3.7, -0.8)
	Limited English Proficiency	3.1	(1.7, 4.4)
	Speech and Language	-1.7	(-2.6, -0.8)
	Mild Mental Retardation	-14.3	(-17.4, -11.3)
	Specific Learning Disability	-7.3	(-8.2, -6.3)
	Other Health Impaired	-7.4	(-8.6, -6.2)
	Special Education - Other	-4.5	(-6.6, -2.3)
	Gifted	10.3	(9.5, 11.2)
	Section 504	-4.0	(-4.8, -3.3)
	Free Lunch	-1.9	(-2.2, -1.6)
	Reduced Price Lunch	-0.9	(-1.3, -0.4)
	Student Absences	-0.3	(-0.3, -0.3)
	Prior Year Mathematics Test	27.9	(27.5, 28.3)
	Prior Year Reading Test	1.2	(1.0, 1.5)
	Prior Year Science Test	5.5	(5.2, 5.8)
	Prior Year Social Studies Test	2.6	(2.3, 2.8)
	Prior Year English-Language Arts Test	2.9	(2.7, 3.2)
Classroom variables	Teacher Absences	0.0	(-0.1, 0.0)
	% Special Education	-7.6	(-12.0, -3.1)
	% Free Lunch	-13.5	(-16.6, -10.4)
	% Reduced Price Lunch	-9.4	(-15.0, -3.7)
	% Minority	-0.2	(-2.5, 2.0)
	% Section 504	3.0	(-3.6, 9.5)
Building variables	% Section 504	16.8	(4.5, 29.1)
	% Free Lunch	16.9	(12.2, 21.5)
	Mean Prior Year Mathematics Test	8.1	(4.2, 12.1)
	Mean Prior Year Science Test	-10.7	(-14.5, -6.9)
	Mean Prior Year Reading Test	7.4	(2.8, 12.1)

In Mathematics, the largest single contributor to a student's Mathematics achievement *among the achievement predictors* was his or her achievement in that domain the prior year. The coefficient for prior achievement in Mathematics was more than five times the value of any other prior achievement variable's coefficient. After considering previous achievement, in Mathematics, all special education diagnoses were associated with negative coefficients; a student identified with a diagnosis of Mild Mental Retardation would be predicted to score 14.3 points lower than one who was not. Boys score slightly higher than girls when controlling for prior achievement, and gifted students score 10.3 points higher than those who are not. The coefficient for students with the Limited English Proficient designation was negative. In addition, the magnitude of the negative coefficient for a student being African American, 4.8 points, while *accounting for prior achievement, poverty, and disability* should be a concern to educators beyond the consideration of teacher preparation.

The magnitude of the coefficient for student absences may surprise some readers; however, it is important to note that this is the effect for *each* day absent. In other words, a student who was absent 20 days would be predicted to score 6 points lower (20 x -0.3 = -6.0) than one with perfect attendance.

Table 6: *Hierarchical Linear Model for Reading Achievement*

Model Level	**Variables Entered**	**Coefficient**	**(CI)**
Student level variables	Gender (Male)	-2.5	(-2.8, -2.1)
	African American	-3.5	(-4.0, -2.9)
	Limited English Proficiency	-4.3	(-5.8, -2.9)
	Mild Mental Retardation	-20.1	(-24.0, -16.2)
	Other Health Impaired	-8.8	(-10.4, -7.1)
	Speech and Language	-5.2	(-6.3, -4.0)
	Specific Learning Disability	-16.4	(-17.6, -15.3)
	Special Education - Other	-7.6	(-10.4, -4.7)
	Gifted	7.9	(7.0, 8.8)
	Section 504	-6.6	(-7.5, -5.7)
	Free Lunch	-3.5	(-3.9, -3.1)
	Reduced Price Lunch	-1.8	(-2.4, -1.2)
	Student Absences	-0.1	(-0.1, -0.1)
	Prior Year English-Language Arts Test	3.4	(3.1, 3.7)
	Prior Year Mathematics Test	4.7	(4.4, 5.0)
	Prior Year Reading Test	14.0	(13.6, 14.3)
	Prior Year Science Test	10.2	(9.9, 10.6)
	Prior Year Social Studies Test	6.9	(6.6, 7.2)
Classroom variables	% Minority	1.1	(-1.1, 3.3)
	% Special Education	-11.9	(-16.5, -7.2)
	% Section 504	8.7	(3.6, 13.7)
	% Free Lunch	-6.7	(-9.8, -3.6)
Building Variables	% Free Lunch	10.3	(5.3, 15.2)
	Mean Prior Year Reading Test	17.6	(13.1, 22.2)
	Mean Prior Year Science Test	-12.4	(-16.5, -8.3)

The pattern of coefficients for Reading parallels the coefficients for Mathematics and at the student level are in the direction that would be expected based on prior research. The coefficient for males reversed direction from Mathematics, suggesting a relative advantage for girls over boys in Reading. This is consistent with previous years' data. Although among previous achievement variables, prior achievement in Reading was the best predictor of current year Reading, the coefficients for prior achievement domains were less differentiated. The negative loading for African American students when accounting for prior achievement and free lunch status is a source of concern.

Table 7: *Hierarchical Linear Model for English English-Language Arts Achievement*

Model Level	Variables Entered	Coefficient	(CI)
Student level variables	Gender (Male)	-11.4	(-11.8, -11.1)
	African American	2.6	(2.2, 3.0)
	Asian American	4.5	(3.4, 5.7)
	Limited English Proficiency	2.2	(0.9, 3.4)
	Emotionally Disturbed	-10.0	(-14.0, -6.1)
	Speech and Language	-3.6	(-4.6, -2.6)
	Mild Mental Retardation	-28.6	(-32.3, -25)
	Other Health Impaired	-9.8	(-11.4, -8.2)
	Specific Learning Disability	-17.2	(-18.3, -16.1)
	Special Education - Other	-7.3	(-9.8, -4.9)
	Gifted	8.2	(7.3, 9.1)
	Section 504	-8.8	(-9.6, -8.0)
	Free Lunch	-2.7	(-3.0, -2.3)
	Reduced Price Lunch	-1.3	(-1.8, -0.8)
	Student Absences	-0.4	(-0.4, -0.4)
	Prior Year Mathematics Test	7.8	(7.5, 8.1)
	Prior Year Reading Test	5.2	(5.0, 5.5)
	Prior Year Science Test	3.4	(3.1, 3.7)
	Prior Year Social Studies Test	3.9	(3.6, 4.3)
	Prior Year English-Language Arts Test	17.2	(16.7, 17.6)
Classroom variables	Teacher Absences	0.0	(-0.1, 0.0)
	% Section 504	12.1	(6.5, 17.7)
	% Gifted	-7.9	(-11.9, -4.0)
	% Gender (Male)	-3.8	(-7.3, -0.3)
	Mean Prior Year Social Studies Test	2.8	(1.8, 3.8)
Building variables	% Free Lunch	6.6	(2.9, 10.3)
	Mean Prior Year Reading Test	7.4	(3.2, 11.7)
	Mean Prior Year English-Language Arts Test	6.4	(3.1, 9.8)
	Mean Prior Year Science Test	-7.9	(-11.3, -4.6)

The pattern of coefficients for English-Language Arts closely parallels the coefficients for Reading at the student level with a few exceptions. Of the previous years' achievement variables, prior year ELA achievement is the largest single predictor of current year's achievement. One notable difference from Reading is that the negative coefficient for males was a great deal larger for English-Language Arts than it was in Reading. This is consistent with previous years' data. In English-Language Arts, all special education variables load negatively and being gifted loads positively. English-Language Arts is the only content area in which the coefficient for African American students was positive, and this is consistent with previous years' data as well.

Table 8: *Hierarchical Linear Model for Science Achievement*

Model Level	Variables Entered	Coefficient	(CI)
Student level variables	Gender (Male)	3.6	(3.3, 3.9)
	African American	-6.2	(-6.6, -5.7)
	Emotionally Disturbed	-4.0	(-7.2, -0.9)
	Mild Mental Retardation	-12.2	(-14.9, -9.5)
	Other Health Impaired	-4.5	(-5.8, -3.2)
	Specific Learning Disability	-5.8	(-6.8, -4.9)
	Speech and Language	-1.5	(-2.4, -0.6)
	Gifted	5.5	(4.8, 6.3)
	Section 504	-2.0	(-2.9, -1.2)
	Free Lunch	-2.4	(-2.8, -2.1)
	Reduced Price Lunch	-0.8	(-1.3, -0.3)
	Student Absences	-0.2	(-0.2, -0.2)
	Prior Year Mathematics test	8.9	(8.6, 9.1)
	Prior Year Reading Test	8.3	(8.1, 8.6)
	Prior Year Science Test	14.7	(14.3, 15.0)
	Prior Year Social Studies Test	7.4	(7.1, 7.6)
	Prior Year English-Language Arts Test	1.3	(1.0, 1.6)
Classroom variables	Teacher Absences	0.0	(-0.1, 0.0)
	% Free Lunch	-8.6	(-11.8, -5.4)
	% Section 504	1.4	(-5.3, 8.0)
Building Variables	% Free Lunch	11.5	(6.6, 16.4)
	% Section 504	17.7	(6.2, 29.1)
	Mean Prior Year Science Test	7.5	(5.6, 9.3)

The base model for Science achievement shares some features with both the Mathematics model and the Reading model. Similar to the results for Mathematics, gender (male) loaded positively, being identified as gifted loaded positively, and being identified as African American loaded negatively. All special education variables loaded negatively. Similar to Reading, prior achievement in the content, Science in this case, was the strongest predictor among the prior achievement variables, but results were not as starkly differentiated as they were in Mathematics.

Table 9: *Hierarchical Linear Model for Social Studies Achievement*

Model Level	Variables Entered	Coefficient	(CI)
Student level variables	Gender (Male)	3.5	(3.2, 3.8)
	African American	-1.0	(-1.5, -0.5)
	Asian American	5.6	(4.5, 6.7)
	Hispanic American	2.9	(2.0, 3.9)
	Section 504	-2.5	(-3.4, -1.6)
	Mild Mental Retardation	-6.1	(-9.0, -3.2)
	Other Health Impaired	-4.1	(-5.4, -2.8)
	Specific Learning Disability	-5.8	(-6.8, -4.9)
	Gifted	8.6	(7.8, 9.5)
	Student Absences	-0.3	(-0.3, -0.2)
	Free Lunch	-3.8	(-4.2, -3.4)
	Reduced Price Lunch	-2.1	(-2.6, -1.6)
	Prior Year Mathematics Test	4.7	(4.5, 5.0)
	Prior Year Reading Test	9.3	(9.0, 9.5)
	Prior Year Science Test	10.8	(10.4, 11.1)
	Prior Year Social Studies Test	12.6	(12.3, 12.9)
	Prior Year English-Language Arts Test	2.5	(2.2, 2.7)
Classroom variables	Teacher Absences	0.0	(-0.1, 0.0)
	% Section 504	-0.6	(-6.2, 5.1)
	% Free Lunch	-13.2	(-16.5, -10.0)
	% Limited English Proficiency	14.8	(7.4, 22.2)
Building variables	% Section 504	25.3	(13.4, 37.2)
	% Free Lunch	13.1	(7.8, 18.4)
	Mean Prior Year Science Test	-6.5	(-11.2, -1.8)
	Mean Prior Year Social Studies Test	13.3	(8.6, 18.0)

As with the other content areas, prior achievement in the domain was the strongest predictor of current year Social Studies achievement among previous years' achievement scores, but it was not as highly differentiated as some of the other content areas. As with all of the content areas, disability status variables loaded negatively and being gifted was advantageous. In Social Studies, the coefficient for males was positive and the coefficient for African American students was negative.

Summary. Generally, the student level models had much in common across content areas. For all areas, prior achievement in the target content area had the largest coefficient *among prior achievement variables*, with achievement in the other four content areas loading to varying degrees. Having a special education diagnosis was a consistent and strong negative predictor of achievement and in many cases (e.g., Mild Mental Retardation), the effect was large. Student absences and free lunch status exhibited consistent, relatively small coefficients. Among the ethnicity factors, no single variable was consistently statistically significant and always loaded in the same direction. However, status as an African American loaded in all of the models and loaded negatively in four of five models.

IV. Assignment of Teachers to Groups

The operational definition of "new teachers" that was employed in the prior VAA-TPP work of teachers in their first two years of teaching was carried forward in this year. Please see previous reports for a description of the rationale for and data that support that designation (e.g., Noell et al., 2008 and Noell et al., 2007), and see the table that follows for the operational definition of *new teacher*.

Table 10: *Teacher Group Assignment*

Group	Criteria
New teachers	1. Teachers in their first or second year of teaching after completing a teacher preparation program leading to initial certification. 2. Certified to teach in the content area. 3. Completed teacher preparation program within 5 years of starting teaching.
Regularly Certified Teachers	1. All other teachers holding a standard certificate. 2. Certified to teach in the content area assessed.
Other	1. Does not conform to any of the categories above.

All subsequent analyses were based upon this categorization combined with the teachers' preparation program that could lead to teacher certification.

V. VAA of Teacher Preparation

Once the final models for student achievement nested within classrooms and schools were developed, these models were used to assess deviations in students' achievement that were associated with being taught by a new teacher from a particular teacher preparation program. This step was the VAA. TPPs were modeled at the teacher level by a series of codes that represented being a new program completer from a particular TPP. Each type of program was modeled separately for each provider: undergraduate, practitioner, master's degree, and non-master's certification only.

The coefficients for recent graduates of particular programs were modeled on the scale of the current *i*LEAP and *LEAP-21* tests due to their importance in high stakes assessment for promotion in grades 4 and 8 as well as their disproportionate weight in School Performance Scores calculated by the State of Louisiana. The tests for 2007 and 2008 had a mean of approximately 300 and a standard deviation of approximately 50 across content areas and grade levels. The results reported below are the mean expected effects for that TPP in comparison to experienced certified teachers.

Prior work (see Noell et al., 2007), used an examination of the ratio of variance within programs to variance between TPPs to arrive at the rule of reporting results only when data were available for 25 or more teachers from a specific program who met the criteria for new teachers (please see Table 10). That rule was carried forward herein.

The assessment was modified slightly from previous years to account for a new issue that arose due to the redesign of TPPs in Louisiana. As some TPPs produced their first cohort of new teachers from their redesigned program, all of the graduates contributing to this assessment were first year teachers. Given that the negative effect for first year teachers is consistently larger than that for second year teachers, the VAA-TPP was modified as follows to equate all programs for the balance of first and second year teachers. Two additional codes were added to level two of the model. These codes identified first and second year teachers respectively. In effect, these codes statistically removed the effect of being a new teacher from the assessment. To retain comparability with previous reports, this effect was reintroduced by simply subtracting the mean effect for first and second year teachers from the coefficient. As a result, the coefficient provides a TPP estimate that controls for the mixture of first and second year teachers and represents the expected result for a balanced mixture of first and second year teachers.

Combining Data Across Years

Following the analytic strategy developed in the VAA-TPP 2007 report, the three consecutive school years were analyzed jointly. The dependent variable was the target achievement test score. The predictor variables were those variables that were identified during model development for that year. All predictor variables for other years were set to 0 (interacted with year). Common codes for TPPs were used across years allowing extraction of cross year coefficients and standard errors from the pooled data.

Additionally, teachers and schools were modeled independently across years. This specification has both analytic and pragmatic advantages. The analytic advantage of specifying schools as independent across years is that it avoids the sometimes problematic assumption that schools are the same organizational units across years. This is obviously not the case when schools are redistricted, have substantial changes in staff, or have their grade configuration revised. One disadvantage is that the model did not capitalize on the repeated observation of teachers across years. However, no software could be identified at the time that these analyses were completed that would allow for such a complex cross-classification structure at the teacher level and that could also resolve a model with so many variables, individuals, and levels. As a result, a model was adopted that treated schools, teachers, and students as independent observations across years.

Performance Bands for Mathematics, Science, and Social Studies

For the 2007 VAA-TPP report, a series of five performance bands was developed in consultation with the then Commissioner of Higher Education and the Associate Commissioner for Teacher Education Initiatives. These levels were designed to create bands of performance that have some intuitive meaning and may help focus readers on clusters of performance rather than a continuous ranking in which the ordering between near neighbors is much more likely to be the result of measurement error than a meaningful difference. Programs are designated according to content area; thus, it is possible for one program to have 5 different levels if its graduates teach in the 5 different content areas. The performance levels are defined in Table 11.

Table 11: *Performance Levels for Teacher Preparation Programs*

Level 1	Programs whose effect estimate is above the mean effect for experienced teachers by its standard error of measurement or more. These are programs for which there is evidence that new teachers are more effective than experienced teachers, but this is not necessarily a statistically significant difference.
Level 2	Programs whose effect estimate is above the mean effect for new teachers by its standard error of measurement or more. These are programs whose effect is more similar to experienced teachers than new teachers.
Level 3	Programs whose effect estimate is within a standard error of measurement of the mean effect for new teachers. These are programs whose effect is typical of new teachers.
Level 4	Programs whose effect estimate is below the mean effect for new teachers by its standard error of measurement or more. These are programs for which there is evidence that new teachers are less effective than average new teachers, but the difference is not statistically significant.
Level 5	Programs whose effect estimate is statistically significantly below the mean for new teachers.

Tables 12-16 below present the VAA estimates for Mathematics, Reading, English-Language Arts, Science, and Social Studies. A more liberal 68% confidence interval (CI) was adopted for this report based on the assumption that for a formative assessment such as this, the consequences of failing to identify an exemplary program or one that is struggling are substantial.

Table 12a: *Undergraduate Teacher Preparation Program Coefficients for **Post-Redesign Programs**: Mathematics*

Level	Teacher Preparation Program	2006-2008 Estimate (CI)	Teachers
3	Louisiana State University Undergraduate	-2.5 (-4.3, -0.7)	25
3	University of Louisiana – Lafayette Undergraduate	-4.3 (-6.4, -2.2)	39

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -2.7.

Table 12b: *Alternate Teacher Preparation Program Coefficients for **Post-Redesign Programs**: Mathematics*

Level	Teacher Preparation Program	2006-2008 Estimate (CI)	Teachers
1	New Teacher Project Practitioner TPP	5.7 (4.0, 7.4)	55
3	Northwestern Louisiana University Practitioner TPP	-0.2 (-3.2, 2.8)	47
3	University of Louisiana – Monroe Master's Alt. Cert.	-1.0 (-2.9, 0.9)	46
3	University of Louisiana – Lafayette NM/CO	-2.2 (-3.9, -0.6)	59
3	Louisiana Resource Center for Educators Practitioner TPP	-2.9 (-4.6, -1.2)	47
3	Louisiana College Practitioner TPP	-3.4 (-5.4, -1.4)	43

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -2.7.

Table 13a: *Undergraduate Teacher Preparation Program Coefficients for* ***Post-Redesign Programs****: Reading*

Level	Teacher Preparation Program	2006-2008 Estimate (CI)	Teachers
3	University of Louisiana - Lafayette Undergraduate	-2.8 (-4.1, -1.5)	49

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -2.8.

Table 13b: *Alternate Teacher Preparation Program Coefficients for* ***Post-Redesign Programs****: Reading*

Level	Teacher Preparation Program	2006-2008 Estimate (CI)	Teachers
1	New Teacher Project Practitioner TPP	4.1 (1.2, 7.0)	30
2	Louisiana College Practitioner TPP	1.2 (-0.9, 3.5)	35
2	Northwestern State University Practitioner TPP	0.4 (-1.8, 2.6)	44
2	University of Louisiana - Monroe Master's Alt. Cert.	0.2 (-2.3, 2.7)	30
3	University of Louisiana - Lafayette NM/CO	-2.9 (-4.8, -0.9)	58
4	Louisiana Resource Center for Educators Practitioner TPP	-6.3 (-8.3, -4.3)	35

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -2.8.

Table 14a: *Undergraduate Teacher Preparation Program Coefficients for **Post-Redesign Programs**: English-Language Arts*

Level	Teacher Preparation Program	2006-2008 Estimate (CI)	Teachers
3	Louisiana State University Undergraduate	-3.7 (-6.0, -1.4)	25
4	University of Louisiana - Lafayette Undergraduate	-4.7 (-6.2, -3.2)	68

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -2.9.

Table 14b: *Alternate Teacher Preparation Program Coefficients for **Post-Redesign Programs**: English-Language Arts*

Level	Teacher Preparation Program	2006-2008 Estimate (CI)	Teachers
2	University of Louisiana - Monroe Master's Alt. Cert.	2.6 (-0.2, 5.4)	41
2	Louisiana State University - Shreveport NM/CO	2.4 (-0.8, 5.6)	26
2	New Teacher Project Practitioner TPP	2.0 (-0.4, 4.4)	41
2	Southeastern Louisiana University Master's Alt. Cert.	1.9 (-0.7, 4.5)	25
2	Louisiana College Practitioner TPP	1.6 (-0.9, 4.1)	36
2	Northwestern State University Practitioner TPP	-0.4 (-2.6, 1.8)	46
3	Louisiana Resource Center for Educators Practitioner TPP	-2.7 (-4.4, -1.0)	43
4	University of Louisiana - Lafayette NM/CO	-4.9 (-6.7, -3.1)	69

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -2.9.

Table 15a: *Undergraduate Teacher Preparation Program Coefficients for **Post-Redesign Programs**: Science*

Level	**Teacher Preparation Program**	**2005-2007 Estimate** (CI)	**Teachers**
3	University of Louisiana – Lafayette Undergraduate	-0.8 (-2.5, 0.9)	39

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -1.4.

Table 15b: *Alternate Teacher Preparation Program Coefficients for **Post-Redesign Programs**: Science*

Level	**Teacher Preparation Program**	**2005-2007 Estimate** (CI)	**Teachers**
1	Northwestern Louisiana University Practitioner TPP	3.7 (1.8, 5.6)	28
1	University of Louisiana - Monroe Master's Alt. Cert.	2.2 (0.0, 4.4)	39
2	New Teacher Project Practitioner TPP	0.9 (-1.3, 3.1)	37
3	Louisiana College Practitioner TPP	-1.4 (-3.7, 0.9)	29
3	Louisiana Resource Center for Educators Practitioner TPP	-1.8 (-3.2, -0.4)	43
3	University of Louisiana – Lafayette NM/CO	-3.1 (-5.2, -1.0)	43

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -1.4.

Table 16a: *Undergraduate Teacher Preparation Program Coefficients for **Post-Redesign Programs**: Social Studies*

Level	Teacher Preparation Program	2006-2008 Estimate (CI)	Teachers
3	University of Louisiana – Lafayette Undergraduate	-3.1 (-4.6, -1.6)	55

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -2.1.

Table 16b: *Alternate Teacher Preparation Program Coefficients for **Post-Redesign Programs**: Social Studies*

Level	Teacher Preparation Program	2006-2008 Estimate (CI)	Teachers
2	University of Louisiana – Monroe Master's Alt. Cert.	1.4 (-0.6, 3.4)	42
2	Northwestern Louisiana University Practitioner TPP	-0.4 (-2.0, 1.2)	29
3	Louisiana College Practitioner TPP	-2.8 (-5.4, -0.2)	40
3	University of Louisiana – Lafayette NM/CO	-2.8 (-5.3, -0.3)	50
3	Louisiana Resource Center for Educators Practitioner TPP	-3.0 (-5.3, -0.7)	28
3	New Teacher Project Practitioner TPP	-3.1 (-6.2, 0.1)	32

Note. The top number in the estimate cells is the mean adjustment to student outcome that would be expected based upon a standard deviation of 50. The numbers in parentheses are the 68% confidence intervals. The mean new teacher effect was -2.1.

Summary of Teacher Preparation Programs' Effects

Estimates for TPPs were generally consistent from the 2008 report to this year's report. For those programs by content area combination represented in both last year's report and this report, 76% of programs fell at the same effectiveness level. In all cases where the level changed, it changed by one level.

Some interesting consistencies emerged in examining the results across the TPPs represented in this year's data. The New Teacher Project, the Master's Program at the University of Louisiana at Monroe, and the Northwestern State University Practitioner Preparation programs had generally positive results. These programs exhibited consistent performance at Level 1 or Level 2 across the five content areas with three Level 3 performances among them (NTP in Social Studies and ULM and NLU in Mathematics). All three of these programs are producing teachers who in aggregate appear to be making a positive contribution to student achievement from the point of entering the classroom. In contrast, the results for both the undergraduate and Non-Master's Certification Only (NM/CO) programs at the University of Louisiana at Lafayette in the domain of English Language Arts fell at level 4. Similarly, the result for the Louisiana Resource Center for Educators (LRCE) in reading was at a level 4. It is important to note that results below level 3 for the ULL-NM/CO and LRCE are consistent with results from last year's assessment and that it is not possible for actions that the faculty at those two institutions took in the last academic year to strengthen their programs to have impacted assessment results yet. Insufficient time has passed for a cohort to complete training under revised procedures and complete an academic year as a teacher.

LSU-Shreveport's non-Master's certification only program and Southeastern's Master's alternate certification program had data sufficient for only one program estimate. Both were in ELA, and both were in Level 2. The Louisiana College Practitioner TPP teachers yielded estimates that ranged from Level 2 (ELA and Reading) to Level 3 (Mathematics, Science, Social Studies).

Among redesigned undergraduate programs, only Louisiana State University and University of Louisiana—Lafayette had data sufficient to generate estimates. LSU's two estimates were at Level 3 (ELA and Mathematics), and ULL's were at Level 3 (Reading, Mathematics, Science, and Social Studies) and Level 4 (ELA).

VI. Teacher Certification and New Teacher Effects

The research team also examined the relationship between teacher effectiveness and teacher certification. For purposes of this analysis, all teachers who were uncertified, teaching on a temporary authority, or were teaching outside their area of certification were pooled. The coefficients in Table 17 demonstrate that teachers who are certified in the content area they are teaching are more effective than those who are not certified to teach that content.

Table 17: *Impact of Teachers who are not Content Certified*

Content	Coefficient (CI)
Mathematics	-2.8 (-3.4, -2.3)
Reading	-2.2 (-2.5, -1.8)
English-Language Arts	-3.6 (-4.2, -3.1)
Science	-1.6 (-2.0, -1.3)
Social Studies	-2.5 (-3.0, -1.9)

Table note. The top value in each cell is the coefficient for that content area. The bottom value in the bottom of the cell is the 95% confidence interval based on the SEM.

VII. Summary

Analyses were conducted to replicate and extend prior statewide analyses for teachers who generally completed their training during the 2004-2005, 2005-2006, and 2006-2007 school years. Construction of the longitudinal database suggested that a sufficient quantity and quality of data are available to support longitudinal analysis of educational inputs such as teacher preparation. For example, the better than 90% linkage rate for student data across years is very encouraging. The proportion of usable records was further decreased by issues such as student mobility and retention, but was above 75% of test takers in all content areas. It is important to acknowledge that as a result of screening measures used with the data, these assessments are for teachers who remain in one school for the academic year, teach the group of students who were promoted the prior year and who remain in that school the entire year. Although this approach selectively excludes teachers and students, it does permit comparison of TPPs in a common database.

The following points are primary findings of each stage of the analyses.

1. The mixed linear models developed for each of the content areas shared a great deal in common. Prior achievement, special education disability status, Section 504 entitlement, receipt of free/reduced price lunch, giftedness, gender, and student absences consistently entered the equations. Being African American was the only ethnicity code that consistently entered models and it loaded negatively in four of five content areas.

2. Some consistency in TPP effects within certification programs was evident with programs exhibiting some clustering in similar places within the distribution of programs. There was a modest amount of variability of coefficients across content areas for the same program. Results were generally consistent at the level of performance bands with the previous report.

3. Examination of the impact of teacher preparation as indexed by certification found that teachers who were not content certified were less effective than content area certified teachers. This difference was particularly large for Reading, English-Language Arts, Mathematics, and Social Studies.

In summary, the data suggest that differences in TPP effectiveness are detectable using data pooled across multiple school years. As the redesigned undergraduate programs produce more teachers employed in Louisiana schools, the number of TPPs whose effectiveness on student learning is possible to estimate will increase.

References

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29,* 37-65.

Goldhaber, D. & Brewer, D. (1997). Why don't schools and teachers seem to matter? Measuring the impact of unobservables on educational productivity. *Journal of Human Resources, 32,* 505-523.

Hill, H. C., Rowan, B. & Ball, Lowenberg, D. L. (2005). The effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42,* 371-406.

Hong, G., & Raudenbush, S. W. (2008). Casual inference for time-varying instructional treatments. *Journal of Educational and Behavioral Statistics, 33,* 333-362.

Martineau, J. A., Subedi, D. R., Ward, K. H., Li, T., Diao, Q., Drake, S., Kao, S. Li, X., Lu, Y., Pang, F., Song, T., & Zheng, Y. (2007). Non-linear unidimensional scale trajectories through multidimensional content spaces: A critical examination of the common psychometric claims of unidimensionality, linearity, and interval-level measurement. in R. W. Lissitz (Ed.) *Assessing and modeling cognitive development in school: Intellectual growth and standard setting.* JAM Press, Maple Grove, MN.

McCaffrey, D. F., Lockwood, J. R., Kortez, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: RAND corporation.

McCaffrey, D. F., Lockwood, J. R., Kortez, D. M., Louis, T. A., & Hamilton, L. S. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29,* 67-102.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* New York: Wiley.

Noell, G. H. (2005). *Assessing teacher preparation program effectiveness: A pilot examination of value added approaches II.* Retrieved from http://asa.regents.state.la.us/TE/value_added_model.

Noell, G. H. (2006). *Annual report of: Value added assessment of teacher preparation.* Retrieved from http://asa.regents.state.la.us/TE/value_added_model.

Noell, G. H., Porter, B. A., & Patt, R. M. (2007). Value added assessment of teacher preparation in Louisiana: 2004-2007. http://www.regents.state.la.us/Academic/TE/Value%20Added.htm.

Noell, G. H., Porter, B. A., Patt, R. M., & Dahir, A. (2008). Value added assessment of teacher preparation in Louisiana: 2004-2005 to 2006-2007. http://www.regents.state.la.us/Academic/TE/2008/Final%20Value-Added%20Report%20(12.02.08).pdf.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*, 237-257.

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

Reckase, M. D. (2004). The real world is more complicated than we would like. *Journal of Educational and Behavioral Statistics, 29,* 117-120.

Sanders, W. L., & Horn, S P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12,* 247-256.

Seltzer, M. H., Frank, K. A., & Bryk, A. S. (1994). The metric matters: The sensitivity of conclusions about growth in student achievement to choice of metric. *Educational Evaluation and Policy Analysis, 16,* 41-49.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal, 113,* 3-33.

Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M. E., Roth, J., Ariet, M., Fisher, T., & Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29,* 11-37.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains. *Review of Educational Research, 73,* 89-122.