**Value Added Assessment of Teacher Preparation in Louisiana:
2005-2006 to 2008-2009**

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

George H. Noell, Ph.D.
R. Maria Knox, M.A.
Michael J. Schafer, B.A.
Department of Psychology

Louisiana State University

August 24, 2010

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, 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 TR-08/10-VAA-94 from the Louisiana Board of Regents.

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, 2007-2008, and 2008-2009. 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, although many of them are, all of Louisiana's teacher preparation programs are not represented in the report. As redesigned programs continue to operate and produce new graduates, the number of programs represented in subsequent reports will increase. 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 2009 report. TPPs generally fell within the same performance bands across this report and the 2009 report within individual content areas. Sixty-five TPPs by content area had sufficient data to be included in this report; last year, 39 TPPs by content area had sufficient data.

Technical Report:
Value Added Assessment of Teacher Preparation in Louisiana:
2005-2006 to 2008-2009

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, 2007-2008, and 2008-2009. These analyses build upon results reported previously in Noell (2006), Noell, Porter, and Patt (2007), Noell, Porter, Patt, and Dahir (2008), and Noell, Gansle, Patt, & Schafer (2009). The VAA-TPP project is a program evaluation study housed in the Department of Psychology at Louisiana State University. The VAA-TPP continues to build longitudinal databases linking students across years and linking those students to their teachers in core content areas in assessed grades. 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 on measures of student achievement. The evaluation team does not have data sufficient to permit examination of 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 studying the process of teacher preparation in the State.

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 specific content domains (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 complex model that includes the grouping of students within classrooms and classrooms within schools. The model examines the degree to which students achieved more or less than would be predicted for them based on prior achievement, attendance, and demographic factors. This information is used to estimate the degree to which students who are taught by new teachers from individual preparation programs achieve at higher or lower levels than their prior achievement history would suggest.

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; Noell et al., 2009). 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 nest students within teachers and teachers within schools, and as a result permits 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 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 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 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 vertically aligned educational tests (Martineau et al., 2007; Seltzer, Frank, & Bryk, 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 Martineau 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. This 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 et al., 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 completing the 4th grade assessment in two consecutive years cannot be analyzed jointly with students who take different tests at two consecutive grade levels. The analyses reported here replicate prior work in which an HLM covariate approach to the data was adopted (Noell, 2006; Noell et al., 2007; Noell et al., 2008; Noell et al., 2009).

Programs Included in the Current Report

In order for a program to be included in the assessment it must have trained a sufficient number of new teachers that complete at least one year of teaching in a public school in a tested subject and grade. Based on prior research (Noell et al., 2007), at least 25 teachers from a particular program must be represented in the data set for a given content area to be included in the value added analyses. 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 workforce 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 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 workforce. This year, far more undergraduate programs (3 to 7 per content area) than last year had data sufficient to appear in the value added analyses. This number will increase in the coming years.

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; Noell et al., 2009). 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 2008-2009 were drawn from the standardized test files (*i*LEAP and *LEAP-21*) for spring 2008 and 2009, the Louisiana Educational Accountability Data System (LEADS, formerly Curriculum database) linking students to teachers, and supplemental student databases. For each year's achievement outcome data described in this report, the previous year's achievement data are included among the predictors of current year performance. 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.

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 of 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 to March 15 of the academic year. 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 all achievement measured at the end of that period to one teacher, and no empirically supported way to attribute portions of it to different teachers. 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 2008-2009*

	Mathematics	**Science**
Assessed students grades 4-9 in 2009	300,806	248,218
Matched to 2008 data	279,954 (93.1%)	232,366 (93.6%)
Consecutive grades assessed	260,323 (86.5%)	217,609 (87.7%)
Single primary school of attendance In curriculum database	241,212 (80.2%)	202,749 (81.7%)

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.

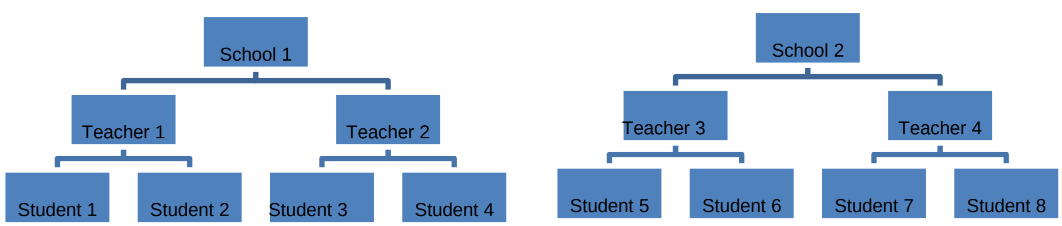
Once students' achievement, demographic, attendance, and course enrollment records were linked, these data were linked to information about their teachers. This included teacher certification data obtained from the Louisiana Department of Education's Division of Planning, Analysis, and Information Resources 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), Noell et al. (2008), and Noell et al. (2009), 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, 2007-2008, and 2008-2009. Please consult previous reports for descriptions of model development. 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
Limited English Proficiency
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

The approach described in previous reports was replicated to determine which of the above variables would be retained at the student, teacher, and school levels for each content area prior to the consideration of teacher preparation program 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 prior work in this area.

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, ∑ 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

Coefficients for variables retained in the model for each year are scaled to the approximate standard deviation of the educational assessments (*i*LEAP and *LEAP*) used in Louisiana: 50. There is general consistency in which variables have been retained in each content area across testing years. The previous year's achievement for a student in a given content has been the strongest predictor of the current year's achievement among prior achievement scores. Having a special education diagnosis is a consistent negative predictor of achievement and in some cases (e.g., Mild Mental Retardation), the effect is large. Interested readers are referred to previous reports for descriptions of the base models from 2005-2006 to

2007-2008 (Noell et al., 2007; Noell et al., 2008; Noell et al., 2009) or to the authors for data from 2008-2009.

IV. Assignment of Teachers to Groups

The operational definition of "new teachers" that was developed in the prior VAA-TPP work was retained. 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 5: *Teacher Group Assignment*

Group	Criteria
New teachers	Teachers who: 1. Were in their first or second year of teaching after completing a teacher preparation program leading to initial certification, 2. Were 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 or 2. Teachers who were certified to teach in the content area assessed.
Other	1. Teachers who do 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 Value Added Assessment (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. These tests have 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.

The assessment was modified slightly from early assessments (prior to 2007-2008) to account for an 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, a large proportion of the graduates contributing to the 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. That convention has been continued from the 2009 report. 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 TPP 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, and to increase the number of programs included in the analyses, four consecutive school years were analyzed jointly and are presented below. 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.

Performance Bands for Mathematics, English-Language Arts, Reading, Science, and Social Studies

Early in this work, a series of five performance level bands was developed in consultation with the then Commissioner of Higher Education and the Associate Commissioner for Teacher Education Initiatives. The performance levels were designed to provide anchors for the numeric representation of the programs by content area. These bands may help readers focus 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. Performance is designated according to content area; thus, it is possible for one program to have 5 different levels if its graduates teach in 5 different content areas. The performance levels are defined in Table 6.

Table 6: *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 7 through 16 present the VAA estimates for Mathematics, English-Language Arts, Reading, Science, and Social Studies. A 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. The tables present programs in sequential order of the magnitude of their current year's effect estimate at the undergraduate or alternate certification program in each content area.

Table 7: *Undergraduate Teacher Preparation Program Coefficients for Post-Redesign Programs in Mathematics*

Level	Program	Effect Estimate 2006-2009	N
2	Southeastern Louisiana University Undergraduate	0.9 (-0.9, 2.7)	28
3	University of New Orleans Undergraduate	-2.1 (-3.6, -0.6)	26
3	Louisiana State University Undergraduate	-2.1 (-3.3, -0.9)	66
3	Louisiana Tech University Undergraduate	-2.7 (-4.1, -1.3)	26
3	University of Louisiana Lafayette Undergraduate	-3.6 (-4.7, -2.5)	110
3	Louisiana State University - Shreveport Undergraduate	-3.6 (-5.1, -2.1)	31
3	McNeese State University Undergraduate	-4.2(-6.3, -2.1)	26

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 -3.1.

Table 8: *Alternate Certification Teacher Preparation Program Coefficients for Post-Redesign Programs in Mathematics*

Level	Program	Effect Estimate 2006-2009	N
1	The New Teacher Project Practitioner TPP	5.1 (4, 6.2)	107
1	Louisiana State University - Shreveport NM/CO	3.4 (0.9, 5.9)	29
2	Southeastern Louisiana University Master's Alternate Certification	2.1 (-1.1, 5.3)	25
3	Northwestern State University of Louisiana Practitioner TPP	-1.5 (-4.2, 1.2)	54
3	University of Louisiana at Monroe Master's Alternate Certification	-2.2 (-4.1, -0.3)	52
3	Louisiana College Practitioner TPP	-2.6 (-4.4, -0.8)	62
3	University of Louisiana Lafayette NM/CO	-3.1 (-4.5, -1.7)	91
3	Louisiana Resource Center for Educators Practitioner TPP	-3.2 (-4.6, -1.8)	63

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 -3.1.

Table 9: *Undergraduate Teacher Preparation Program Coefficients for Post-Redesign Programs in English-Language Arts*

Level	Program	Effect Estimate 2006-2009	N
3	Northwestern State University of Louisiana Undergraduate	-2.8 (-4.6, -1)	35
3	Louisiana State University - Shreveport Undergraduate	-3.0 (-5, -1)	37
3	McNeese State University Undergraduate	-3.1 (-5.5, -0.7)	25
3	Louisiana State University Undergraduate	-3.6 (-4.9, -2.3)	68
3	Southeastern Louisiana University Undergraduate	-3.9 (-5.2, -2.6)	42
4	University of Louisiana Lafayette Undergraduate	-4.4 (-5.6, -3.2)	124

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 10: *Alternate Certification Teacher Preparation Program Coefficients for Post-Redesign Programs in English-Language Arts*

Level	Program	Effect Estimate 2006-2009	N
2	Louisiana State University - Shreveport NM/CO	1.8 (-1.1, 4.7)	37
1	The New Teacher Project Practitioner TPP	1.7 (0.1, 3.3)	77
2	Southeastern Louisiana University Master's Alternate Certification	1.6 (-0.7, 3.9)	41
2	Louisiana College Practitioner TPP	1.5 (-0.8, 3.8)	51
2	University of Louisiana at Monroe Master's Alternate Certification	0.8 (-1.7, 3.3)	48
2	Northwestern State University of Louisiana Practitioner TPP	-0.2 (-2.3, 1.9)	49
3	Louisiana Resource Center for Educators Practitioner TPP	-2.9 (-4.5, -1.3)	54
4	University of Louisiana Lafayette NM/CO	-5.1 (-6.7, -3.5)	89

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 11: *Undergraduate Teacher Preparation Program Coefficients for Post-Redesign Programs in Reading*

Level	Program	Effect Estimate 2006-2009	N
3	Louisiana State University Undergraduate	-2.2 (-3.6, -0.8)	46
3	University of Louisiana Lafayette Undergraduate	-3.2 (-4.2, -2.2)	99
3	Louisiana State University - Shreveport Undergraduate	-4.1 (-6.1, -2.1)	30

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 -2.6.

Table 12: *Alternate Certification Teacher Preparation Program Coefficients for Post-Redesign Programs in Reading*

Level	Program	Effect Estimate 2006-2009	N
1	The New Teacher Project Practitioner TPP	2.5 (0.6, 4.4)	51
1	Louisiana College Practitioner TPP	1.8 (0.1, 3.5)	56
2	Southeastern Louisiana University Master's Alternate Certification	1.6 (-1, 4.2)	25
2	Louisiana State University - Shreveport NM/CO	1.1 (-1.8, 4)	28
2	Northwestern State University of Louisiana Practitioner TPP	-0.1 (-2.3, 2.1)	47
3	University of Louisiana at Monroe Master's Alternate Certification	-0.7 (-2.8, 1.4)	37
3	University of Louisiana Lafayette NM/CO	-2.9 (-4.6, -1.2)	78
4	Louisiana Resource Center for Educators Practitioner TPP	-5.0 (-6.7, -3.3)	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 -2.6.

The LRCE Practitioner TPP first received a level 5 result in the 2008 report. As a result of that feedback and their self assessment curricular changes were made that year. Assuming that they were immediately brought to scale and were successful, the first year that the changes should be evident in this report would be the 2011 report (see the time line above). However, the authors examined the data for just the last two years for LRCE to ascertain whether any trend was evident. Examining just the last two years of data, the results for LRCE would move up to -1.8 (SEM 2.2). Although this would be consistent with a Level 3 result, it is important to recognize two important limitations to these data. First, it is based on only 13 teachers which is below the number set as a standard for reporting results. Second, it is based on a different time frame than the other results reported herein and is as a result is not directly comparable. Acknowledging those limitations, the data do suggest a positive trend for LRCE.

Table 13: *Undergraduate Teacher Preparation Program Coefficients for Post-Redesign Programs in Science*

Level	Program	Effect Estimate 2006-2009	N
1	Louisiana State University Undergraduate	1.1 (0.1, 2.1)	50
2	Southeastern Louisiana University Undergraduate	0.6 (-0.9, 2.1)	29
3	University of Louisiana Lafayette Undergraduate	-2 (-3, -1)	106

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 -1.5.

Table 14: *Alternate Certification Teacher Preparation Program Coefficients for Post-Redesign Programs in Science*

Level	Program	Effect Estimate 2006-2009	N
1	Northwestern State University of Louisiana Practitioner TPP	3.3 (1.4, 5.2)	31
1	Southeastern Louisiana University Master's Alternate Certification	2.6 (0.7, 4.5)	36
2	Louisiana State University - Shreveport NM/CO	2.5 (-0.4, 5.4)	25
1	The New Teacher Project Practitioner TPP	2.1 (0.5, 3.7)	73
2	University of Louisiana at Monroe Master's Alternate Certification	0.8 (-1.2, 2.8)	47
3	Louisiana Tech University NM/CO	-0.6 (-2.3, 1.1)	25
3	Louisiana College Practitioner TPP	-0.7 (-2.4, 1)	49
3	Louisiana Resource Center for Educators Practitioner TPP	-1.4 (-2.6, -0.2)	52
4	University of Louisiana Lafayette NM/CO	-3.4 (-5.2, -1.6)	61

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 -1.5.

Table 15: *Undergraduate Teacher Preparation Program Coefficients for Post-Redesign Programs in Social Studies*

Level	Program	Effect Estimate 2006-2009	N
2	Louisiana State University Undergraduate	0.2 (-1.2, 1.6)	58
3	University of New Orleans Undergraduate	-1.0 (-3.4, 1.4)	25
4	University of Louisiana Lafayette Undergraduate	-3.8 (-4.9, -2.7)	110
3	Louisiana State University - Shreveport Undergraduate	-3.9 (-6, -1.8)	32
5	McNeese State University Undergraduate	-5.7 (-7.5, -3.9)	30

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 -2.0.

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

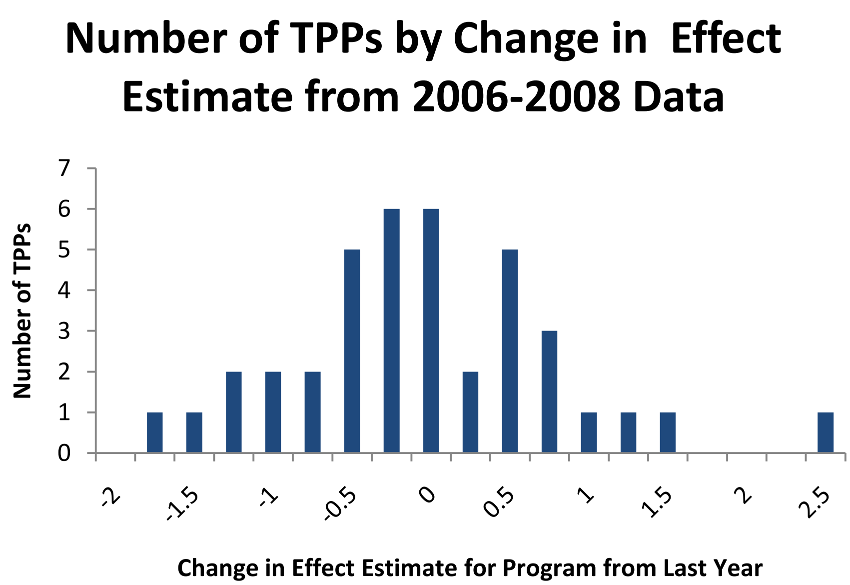
Level	Program	Effect Estimate 2006-2009	N
1	Louisiana State University - Shreveport NM/CO	4.0 (1.5, 6.5)	27
1	Southeastern Louisiana University Master's Alternate Certification	2.6 (0.4, 4.8)	30
1	University of Louisiana at Monroe Master's Alternate Certification	1.9 (0, 3.8)	46
3	Louisiana College Practitioner TPP	-0.4 (-2.5, 1.7)	58
3	Northwestern State University of Louisiana Practitioner TPP	-1.0 (-2.6, 0.6)	33
3	The New Teacher Project Practitioner TPP	-2.6 (-4.7, -0.5)	56
3	University of Louisiana Lafayette NM/CO	-2.8 (-4.9, -0.7)	69
3	Louisiana Resource Center for Educators Practitioner TPP	-3.0 (-4.9, -1.1)	38

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 -2.0.

Stability of Teacher Preparation Programs' Effects

The estimates for TPPs were generally consistent from the previous report. For those programs by content area combinations represented in both last year's report and this report, 77% of programs fell at the same level. In all cases where the level changed, it changed by one level. Ten percent of programs (4) increased levels; 13% of programs (5) decreased levels. Changes to this year's estimate from last year's estimate ranged between -1.8 points and +2.4 points. The distribution of changes is depicted in Figure 2. A high level of stability is to be expected given the extent to which the assessments across the last two years' reports share data.

Figure 2. *Distribution of Changes in Effect Estimates*



There were three programs whose results were generally consistent with the student achievement results of experienced certified teachers. The Louisiana State University - Shreveport NM/CO program and Southeastern Louisiana University Master's Alternate Certification program both had two content areas at level 1 and three contents at level 2. The New Teacher Project Practitioner Program had four content areas at level 1 and one content area at level 3. These programs are producing teachers who in aggregate appear to be making a positive contribution to student achievement from the time they complete their training program and begin teaching.

There were five programs whose effect estimates were a mixture of levels 1, 2, and 3. The Louisiana College Practitioner TPP, the Louisiana State University Undergraduate program, the Northwestern State University of Louisiana Practitioner TPP program, the University of Louisiana at Monroe Master's Alternate Certification, and the Southeastern Louisiana University Undergraduate program obtained results across content areas that varied between the typical results for new teachers and those of veteran teachers.

There were four programs whose new teachers performed primarily at the same level as average teachers. The Louisiana State University - Shreveport Undergraduate program had 4 content areas at level 3; the University of New Orleans Undergraduate program had 2 content areas at level 3; and the Louisiana Tech University NM/CO program, Louisiana Tech University Undergraduate program, and Northwestern State University of Louisiana Undergraduate program had 1 content area at level 3.

There were no programs whose effect estimates were primarily below those of new teachers. However, there were four programs whose effect estimates were at or below the level of new teachers. The bulk of the Louisiana Resource Center for Educators Practitioner TPP, the McNeese State University Undergraduate program, the University of Louisiana Lafayette NM/CO program, and the University of Louisiana Lafayette Undergraduate program's effect estimates were at the level of new teachers but each also had at least one effect estimate that was at level 4 or 5. These programs had no effect estimates that were at or above the level of experienced certified teachers.

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 in which 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	-2.4 (-3.2, -1.6)
English-Language Arts	-3.3 (-4.3, -2.3)
Reading	-2.2 (-2.8, -1.6)
Science	-1.6 (-2.2, -1.0)
Social Studies	-2.4 (-3.2, -1.6)

Table note. The first value in the right column is the coefficient for that content area. The values in parentheses are the 95% confidence interval based on the standard error of measurement.

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, 2006-2007, and 2007-2008 school years. It is important to acknowledge that as a result of screening measures used with the data, these assessments describe only teachers who remain in one school for the academic year and teach the group of students who remained with them from the beginning of the school year until testing time and who were promoted the prior year. Although this approach selectively excludes teachers and students, it does permit comparison of TPPs in a common database.

1. Some consistency in TPP effects within certification programs continues to be 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.

2. One result that is relatively obvious is that the alternative TPP had more positive results (Levels 1 and 2) than the undergraduate programs. They were similar in the number of Level 4 and 5 programs. This finding will naturally raise the issue of why this would be the case. It is possible that alternative programs provide an intensity of practical training that better prepares new teachers than undergraduate routes. It is also possible that they are drawing from a different population of potential teachers (more experienced candidates for whom teaching is not a first fulltime or professional position), who are more ready to be successful in the classroom when they complete their program. It is also possible that alternative routes, notably the practitioner route, simply provide more experience teaching prior to a candidate completing their course of preparation. The authors will work with the research team lead by Dr. Jeanne Burns to plan a series of investigations to examine these hypotheses, as well as others that may emerge.

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 analyses suggest that differences in TPP effectiveness for the 59 programs examined with 25 or more new teachers in a content area are detectable using data that have been 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 continue to increase.

References

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). *Technical report of: Assessing teacher preparation program effectiveness: A pilot examination of value added approaches (2005).* Retrieved from the Louisiana Board of Regents website: http://regents.louisiana.gov/Academic/TE/technical_report_200405.pdf.

Noell, G. H. (2006). *Technical report of: Assessing teacher preparation program effectiveness: A pilot examination of value added approaches (2006).* Retrieved from the Louisiana Board of Regents website: http://regents.louisiana.gov/Academic/TE/technical_report.pdf.

Noell, G. H., Gansle, K. A., Patt, R. M., & Schafer, M. J. (2009). *Value added assessment of teacher preparation in Louisiana: 2005-2006 to 2007-2008.* Retrieved from the Louisiana Board of Regents website: http://regents.louisiana.gov/Academic/TE/2009/2008-09VATechnical(8.24.09).pdf.

Noell, G. H., Porter, B. A., & Patt, R. M. (2007). *Value added assessment of teacher preparation in Louisiana: 2004-2006.* Retrieved from the Louisiana Board of Regents website: http://regents.louisiana.gov/Academic/TE/2007/VAA%20TPP%20Technical%20Report%2010-24-2007.pdf.

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.* Retrieved from the Louisiana Board of Regents website: 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.